Our Ref. No.: HYS-46CIP

METHODS AND MATERIALS RELATING TO NOVEL C1q DOMAIN-CONTAINING POLYPEPTIDES AND POLYNUCLEOTIDES

Express Mail Label No.: EV 345 137 347 US

METHODS AND MATERIALS RELATING TO NOVEL C1q DOMAIN-CONTAINING POLYPEPTIDES AND POLYNUCLEOTIDES

This application is a continuation-in-part application of PCT Application Serial No. PCT/US02/38526 filed December 3, 2003, entitled "Methods and Materials Relating to Novel Polypeptides and Polynucleotides," Attorney Docket No. HYS-B1CIP/PCT, which is a continuation-in-part application of U.S. Application Serial No. 10/005,499 filed December 3, 2001 entitled "Methods and Materials Relating to Novel Secreted C1q domain-containing Polypeptides and Polynucleotides," Attorney Docket No. HYS-46, which in turn is a continuation-in part application of U.S. Application Serial No. 10/296,115 (I.A. filing date of December 22, 2000) filed on June 24, 2003 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 784CIP3A/US, which is a national phase application of PCT Application Serial No. PCT/US00/35017 filed December 22, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 784CIP3A/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/552,317 filed April 25, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 784CIP (now abandoned), which in turn is a continuation-in-part application of U.S. Application Serial No. 09/488,725 filed January 21, 2000 entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 784; U.S. Application Serial No. 10/286,897 filed November 1, 2002, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 784CIP4, which is a continuation-in-part application of U.S. Application Serial No. 10/258,898 (I.A. filing date of December 22, 2000) filed on July 21, 2003 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 784CIP2-2F/US, which is a national phase application of PCT application Serial No. PCT/US00/34263, filed December 22, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 784CIP2-2F/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/620,312 (now U.S. Patent No. 6,569,662) filed July 19, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 784CIP2B; U.S. Application Serial No. 10/276,774 (I.A. filing date of February 5, 2001) filed on June 24, 2003 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 787CIP3/US, which is a national phase application of PCT Application Serial No. PCT/US01/03800 filed February 5, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 787CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/560,875 filed April 27, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 787CIP, which in turn is a continuation-in-part application of

U.S. Application Serial No. 09/496,914 filed February 3, 2000 entitled "Novel Contigs Obtained from Various Libraries," Attorney Docket No. 787 (now abandoned); U.S. Application Serial No. 10/293,244 filed November 12, 2002, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 787CIP4A, which in turn is a continuation-in-part application of U.S. Application Serial No. 10/258,899 (I.A. filing date of February 5, 2001) entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 787CIP2-2G/US, which in turn is a national phase application of PCT application Serial No. PCT/US01/04098, filed February 5, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 787CIP2-2G/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/598,075 filed June 20, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 787CIP2G (now abandoned); U.S. Application Serial No. 10/450,763 (I.A. filing date of March 30, 2001) entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 790CIP3/US, which in turn is a national phase application of PCT Application Serial No. PCT/US01/08631 filed March 30, 2001 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 790CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/649,167 filed August 23, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 790CIP (now abandoned), which in turn is a continuation-in-part application of U.S. Application Serial No. 09/540,217 filed March 31, 2000 entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 790 (now abandoned); U.S. Application Serial No. 10/416,991 (I.A. filing date of November 30, 2001) entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 799CIP/US, which is a national phase application of PCT Application Serial No. PCT/US01/47004 filed November 30, 2001, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 799CIP/PCT, which in turn is a continuation-in-part application of U.S. application Serial No. 09/728,952 filed November 30, 2000 entitled "Novel Nucleic Acids and Polypeptides", Attorney Docket No. 799 (now abandoned); PCT Application Serial No. PCT/US02/22858 filed July 19, 2002, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 805A/PCT, which claims the benefit of priority of U.S. Provisional application Serial No. 60/306,971 filed July 21, 2001 entitled "Novel Nucleic Acids and Polypeptides", Attorney Docket No. 805 (now expired); PCT Application Serial No. PCT/US02/29636 filed September 18, 2002, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 808ACIP/PCT, which claims the benefit of priority to U.S. Provisional Application 60/323,349 filed September 18, 2001, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 808 (now expired); PCT Application Serial No. PCT/US02/29964 filed September 19, 2002, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 809ACIP/PCT, which claims the benefit of priority to U.S. Provisional

Application Serial No. 60/323,739 filed July 21, 2001, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 809 (now expired); and PCT Application Serial No. PCT/US02/30474 filed September 24, 2002, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 810CIP/PCT, which claims the benefit of priority to U.S. Provisional Application Serial No. 60/324,631 filed September 24, 2001, entitled "Novel Nucleic Acids and Polypeptides," Attorney Docket No. 810 (now expired); all of which are herein incorporated by reference in their entirety.

1. BACKGROUND

1.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods. In particular, the invention relates to C1q domain-containing polypeptides and polynucleotides and uses thereof.

1.2 BACKGROUND ART

The complement pathway is one of the major effector mechanisms of humoral immunity as well as an important mechanism of innate immunity. One of the main functions of proteins involved in the complement pathway is in microbial cell lysis. The products of complement activation become covalently attached to microbial cell surfaces or to antibodies bound to microbes and other antigens (reviewed in Abbas *et al., Cellular and Molecular Immunology*, 4th ed., W.B. Saunders Co., Philadelphia, PA, 2000, pp. 316-331, herein incorporated by reference in its entirety). C1q protein is the first subcomponent of the classical complement pathway and binds to antigen-bound antibodies. The C1q subcomponent contains six A, six B, and six C chains (C1qA, B, and C) and forms a bouquet-like structure with six branches (Figure 1). Each of the 6 heads of the C1q bouquet is a heterotrimer of C-terminal globular regions of A, B, and C chains (the C1q domain), whereas the arms are triple helices formed by the collagen-like regions of these three chains.

In recent years, many non-complement proteins have been identified that contain C1q domains. Most of them have a similar structure comprising a leading signal peptide, followed by a collagen-like region, and a C-terminal C1q domain (reviewed in Kishore and Reid, *Immunopharmacology* 42:15-21 (1999), herein incorporated by reference in its entirety). Both the structure and sequence of the C1q domains are conserved; however, the function of these C1q proteins is not conserved. There are many C1q domain containing proteins that are not involved in the complement pathway. These proteins include: human type VIII and type X

collagen (Yamaguchi *et al., J. Biol. Chem.* 270:16022 (1989), Ninomiya *et al., J. Biol. Chem.* 274:16773 (1999), respectively), precerebellins (neuronal proteins) (Urade *et al, Proc. Natl. Acad. Sci. USA* 88:1069 (1991)), chipmunk hibernation proteins (Takamatsu *et al., Mol. Cell. Biol.* 13:1516 (1993)), multimerin (a human endothelial cell protein) (Hayward *et al., J. Biol. Chem.* 270:18246 (1995)), adiponectin (Scherer *et al., J. Biol. Chem.* 270:26746 (1995)), saccular collagen (Davis *et al., Science* 163:1031 (1995)), and EMLIN which is found in elastin-rich tissues (Doliana *et al., J. Biol. Chem.* 274:16773 (1999) this and all other references are herein incorporated by reference in their entirety).

There are four members in the precerebellen family, CBLN1 to 3. Cerebellin is a 16 amino acid neuropeptide that is most abundantly expressed in the cerebellum and has been shown to enhance secretory activity of the adrenal gland (Mazzocchi *et al., J. Clin. Endocrinol. Metab.* 84:632-635 (1999); Albertin *et al., Neuropeptides* 34:7-11 (2000); both of which are herein incorporated by reference in their entirety). Similar to other neuropeptides, cerebellin is derived from a precursor protein named precerebellin 1 (CBLN1) (Urade *et al.,* 1991, *supra,*). Precerebellin 1 is composed of a signal peptide, an N-terminal region, a cerebellin motif, and a C-terminal C1q domain; however, it does not contain a collagen-like region (Urade *et al.*, 1991, *supra*).

The chipmunk hibernation-associated proteins HP-20, 25, 27, and 55 form a 140 kD complex in plasma. The expression level of this complex tightly associates with the hibernation status of the animal: it drops before the onset of hibernation and increases before hibernation ends (Takamatsu *et al., Mol. Cell Biol.* 13:1516-1521 (1993), herein incorporated by reference in its entirety). HP-20, 25, and 27 are homologous to each other and each contains a collagen-like region followed by a C-terminal C1q domain. These genes are present but not expressed in a non-hibernating squirrel (Takamatsu *et al.*, 1993, *supra*).

Short chain collagens include two type VIII collagens, α1 (COL8A1) and α2 (COL8A2), and one type X collagen (COL10A1). Collagen VIII is a major component of Descemet's membrane, the basement membrane of corneal endothelial cells (Yamaguchi *et al., J. Biol. Chem.* 264:16022-16029 (1989), herein incorporated by reference in its entirety), whereas collagen X is specifically expressed by hypertrophic condrocytes during bone development (Thomas *et al, Biochem. Soc. Trans.* 19:804-808 (1991), herein incorporated by reference in its entirety).

Adiponectin (also known as Acrp30, AdipoQ, APM1, and GBP28) is an anti-diabetic hormone exclusively produced by adipose tissue and released into the circulation that regulates glucose and lipid metabolism (reviewed in Pajvani and Scherer, *Curr. Diab. Rep.* 3:207-213

(2003), herein incorporated by reference in its entirety). Specifically, adiponectin stimulates glucose utilization and fatty-acid oxidation by activating the 5'-AMP-activated protein kinase (Yamauchi *et al., Nat. Med.* 8:1288-1295 (2002), herein incorporated by reference in its entirety). Adiponectin knockout mice show delayed clearance of free fatty acid in plasma, a high level of plasma TNF-α, and severe diet-induced insulin resistance (Maeda *et al., Nat. Med.* 8:731-737 (2002), herein incorporated by reference in its entirety). Structurally, adiponectin contains a leading signal peptide, a collagen-like region, and a C-terminal C1q domain. The crystal structure of the C1q domain of adiponectin shows a significant similarity to that of tumor necrosis factor a (TNFa), indicating an evolutionary connection between C1q-related proteins and TNF family members (Shapiro and Scherer, *Curr. Biol.* 8:335-338 (1998), herein incorporated by reference in its entirety).

Discovery and characterization of other C1q related polypeptides will be advantageous to diagnose and treat a variety of disorders, including inflammation, immune disorders, diabetes, and lipid metabolism.

2. SUMMARY OF THE INVENTION

This invention is based on the discovery of novel C1q domain-containing polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors such as expression vectors containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides, and cells genetically engineered to express such polynucleotides.

The compositions of the invention provide isolated polynucleotides that include, but are not limited to, a polynucleotide comprising the nucleotide sequence set forth in SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58; a polynucleotide comprising the full length protein coding sequence of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58); and a polynucleotide comprising the nucleotide sequence of the mature protein coding sequence of any of SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any of the nucleotide sequences set forth in SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58; (b) a nucleotide sequence

encoding any of SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69; a polynucleotide which is an allelic variant of any polynucleotides recited above having at least 70% polynucleotide sequence identity to the polynucleotides; a polynucleotide which encodes a species homolog (*e.g.* orthologs) of any of the peptides recited above; or a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptide comprising SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or unique identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention further provides cloning or expression vectors comprising at least a fragment of the polynucleotides set forth above and host cells or organisms transformed with these expression vectors. Useful vectors include plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The compositions of the present invention include polypeptides comprising, but not limited to, an isolated polypeptide selected from the group comprising the amino acid sequence of SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the protein sequences listed as SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69 and substantial equivalents thereof that retain biological or immunological activity are also contemplated. The polypeptides of the invention may be wholly or

partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (*e.g.* host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Pharmaceutical compositions of the invention may comprise a polypeptide of the invention and an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The invention also relates to methods for producing a polypeptide of the invention comprising culturing host cells comprising an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the protein or peptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such a process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use in an array, use in computer-readable media, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of antisense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath *et al., Science* 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. Furthermore, antibodies, particularly monoclonal antibodies, are useful for binding to and/or inhibiting the function of polypeptides of the invention and therefore may be useful in the treatment of diseases in which the polypeptides are over-expressed or have increased activity.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective

amount of a composition comprising a peptide of the present invention and a pharmaceutically acceptable carrier.

The methods of the invention also provide methods for the treatment of disorders as recited herein which comprise the administration of a therapeutically effective amount of a composition comprising a polynucleotide or polypeptide of the invention and a pharmaceutically acceptable carrier to a mammalian subject exhibiting symptoms or tendencies related to disorders as recited herein. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising the step of administering a composition comprising compounds and other substances that modulate the overall activity of the target gene products and a pharmaceutically acceptable carrier. Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity. Specifically, methods are provided for preventing, treating or ameliorating a medical condition, including viral diseases, which comprises administering to a mammalian subject, including but not limited to humans, a therapeutically effective amount of a composition comprising a polypeptide of the invention or a therapeutically effective amount of a composition comprising a binding partner of (*e.g.*, antibody specifically reactive for) C1q domain-containing polypeptides of the invention. The mechanics of the particular condition or pathology will dictate whether the polypeptides of the invention or binding partners (or inhibitors) of these would be beneficial to the individual in need of treatment.

According to this method, polypeptides of the invention can be administered to produce an *in vitro* or *in vivo* inhibition of cellular function. A polypeptide of the invention can be administered *in vivo* alone or as an adjunct to other therapies. Conversely, protein or other active ingredients of the present invention may be included in formulations of a particular agent to minimize side effects of such an agent.

The invention further provides methods for manufacturing medicaments useful in the above-described methods.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample (e.g., tissue or sample). Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions.

The invention provides a method for detecting a polypeptide of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the

polypeptide under conditions and for a period sufficient to form the complex and detecting formation of the complex, so that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (*i.e.*, increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention.

The invention provides a method for identifying a compound that binds to the polypeptide of the present invention comprising contacting the compound with the polypeptide under conditions and for a time sufficient to form a polypeptide/compound complex and detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide is identified.

Also provided is a method for identifying a compound that binds to the polypeptide comprising contacting the compound with the polypeptide in a cell for a time sufficient to form a polypeptide/compound complex wherein the complex drives expression of a reporter gene sequence in the cell and detecting the complex by detecting reporter gene sequence expression so that if the polypeptide/compound complex is detected a compound that binds to the polypeptide is identified.

3. BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Schematic diagram of the bouquet-like structure of C1q and heterotrimeric assembly of C1q A, B, and C chains.

Figure 2: Schematic diagrams representing domain structures and exon patterns of human C1q domain-containing proteins. Vertical lines indicate exon boundaries.

For All Figures except Figures 16 and 17, amino acids are abbreviated as follows: A=Alanine, C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, and Y=Tyrosine.

Figure 3A-B: Sequence alignment of C1q domain regions of human C1q domain-containing proteins. Conserved residues are boxed whereas highly conserved residues (only different in 4 or less sequences) are shaded. Arrows underneath the alignment represent β-strand positions found in the crystal structure of the adiponectin C1q domain. The C1q domains are from the following CDCP proteins: Adiponectin (SEQ ID NO: 94), AQL1 (SEQ ID NO: 95), AQL2 (SEQ ID NO: 96), C1QA (SEQ ID NO: 97), C1QB (SEQ ID NO: 98), C1QC (SEQ ID NO: 99), C1QTNF1 (SEQ ID NO: 100), C1QTNF2 (SEQ ID NO: 101), C1QTNF3 (SEQ ID NO: 102), C1QTNF4.1 (SEQ ID NO: 103), C1QTNF4.2 (SEQ ID NO: 104), C1QTNF5 (SEQ ID NO: 105), C1QTNF6 (SEQ ID NO: 106), C1QTNF7 (SEQ ID NO: 107), C1QTNF8 (SEQ ID NO: 108), CBLN1 (SEQ ID NO: 109), CBLN2 (SEQ ID NO: 110), CBLN3 (SEQ ID NO: 111), CBLN4 (SEQ ID NO: 112), CRF1 (SEQ ID NO: 113), CRF2 (SEQ ID NO: 114), Gliacolin1 (SEQ ID NO: 115), Gliacolin2 (SEQ ID NO: 116), Otolin (SEQ ID NO: 117), COL8A1 (SEQ ID NO: 118), COL8A2 (SEQ ID NO: 119), COL10A1 (SEQ ID NO: 120), C1QDC1 (SEQ ID NO: 121), EMILIN1 (SEQ ID NO: 122), EMILIN2 (SEQ ID NO: 123), EMILIN3 (SEQ ID NO: 124), and multimerin (SEQ ID NO: 125).

Figure 4: Three-dimensional (3D) structures of adiponectin, AQL1, C1qTNF7 and cortical vesicle protein CV34-23. The crystal structure of adiponectin (accession number 1C28, RCSB Protein Data Bank (Berman *et al., Nucl. Acids Res.* 28:235-242 (2000) herein incorporated by reference in its entirety) and structural models of human AQL1, human C1qTNF7, and sea urchin (Strongylocentrotus purpuratus) cortical vesicle protein CV34-23 based on the structure of adiponectin are shown. All four of the structures follow a ten β-strand jelly-roll folding topology (Shapiro and Scherer, 1998, *supra*). The eight amino acids that are conserved over all human C1q proteins in Figure 3 are labeled. The location of these residues suggests that they may be essential for effective packing of the hydrophobic core of the molecules. Seven of these eight amino acids are conserved in the CV34-23 protein.

Figure 5: Phylogenetic tree of all human C1q domains. A phylogenetic dendrogram (phylogram) was generated from a Clustal-W alignment of all human C1q domain sequences using the TreeTop program (GeneBee Group, Belozersky Institute, Moscow State University, Russia). The branch lengths (x-axis) in the rectangular cladogram represent the distances among those sequences calculated using the BLOSUM62 substitution matrix. The numbers at branching points are bootstrap values indicating the reliability of assignment.

Figure 6: Clustal-W multiple amino acid sequence alignment of SEQ ID NOs: 4, 7, 10, and 19 with human similar-to-ACRP30, gi:29738938 (SEQ ID NO: 70), wherein identical residues are represented by an asterisk (*), conservative substitutions are represented by a colon (:), and semi-conservative substitutions are represented by a period (.).

Figure 7: BLASTP amino acid sequence alignment of SEQ ID NO: 24 with human α1 type VIII collagen precursor, gi:17738302 (SEQ ID NO: 71) showing 99% identity over 744 amino acids of SEQ ID NO: 71. Gaps are represented as dashes.

Figure 8: Sequence alignment of otolins from Fugu [*Takifugu rubripes* (SEQ ID NO: 88)], bluegill sunfish [*Lepomis macrochirus* (SEQ ID NO: 89)], chum salmon [*Oncorhynchus keta* (SEQ ID NO: 90)], human [*Homo sapiens* (SEQ ID NO; 91)], mouse [*Mus musculus* (SEQ ID NO: 92)], and rat [*Rattus norvegicus* (SEQ ID NO: 93)]. Conserved residues are boxed. The C1q domain region is marked by a line on top of the alignment.

Figure 9: BLASTP amino acid sequence alignment of SEQ ID NO: 27 and human similar to otolin-1, gi:22041493 (SEQ ID NO: 78) showing 94% identity over 459 amino acids of SEQ ID NO: 78.

Figure 10: Clustal-W multiple amino acid sequence alignment of SEQ ID NOs: 32, 34, 38, and 41 with murine gliacolin, gi:23680960 (SEQ ID NO: 72), wherein identical residues are represented by an asterisk (*), conservative substitutions are represented by a colon (:), and semi-conservative substitutions are represented by a period (.).

Figure 11: Clustal-W multiple amino acid sequence alignment of SEQ ID NOs: 32, 34, 38, and 41 with human C1q-related factor,gi:5729785 (SEQ ID NO: 73), wherein identical residues are represented by an asterisk (*), conservative substitutions are represented by a colon (:), and semi-conservative substitutions are represented by a period (.).

Figure 12A-B: Clustal-W multiple sequence alignment of SEQ ID NOs: 46, 48, and 51 with human C1q domain-containing 1 isoform L (EEG1L), gi:23503235 (SEQ ID NO: 74), wherein identical residues are represented by an asterisk (*), conservative substitutions are represented by a colon (:), and semi-conservative substitutions are represented by a period (.).

Figure 13: BLASTP amino acid sequence alignment of SEQ ID NO: 55 and human EMILIN-2 precursor, gi:14042988 (SEQ ID NO: 77), showing 98% identity over 267 amino acids of SEQ ID NO: 77, wherein gaps are presented as dashes.

Figure 14: BLASTP amino acid sequence alignment of SEQ ID NO: 59 with human C1qTNF-7 gi:13994280 (SEQ ID NO: 75) showing 100% identity over 289 amino acids of SEQ ID NO: 75, wherein gaps are presented as dashes.

Figure 15: BLASTP amino acid sequence alignment of SEQ ID NO: 63 with human C1qTNF-6 gi:32967294 (SEQ ID NO: 76) showing 100% identity over 259 amino acids of SEQ ID NO: 76, wherein gaps are presented as dashes.

Figure 16A-F: Multiple nucleic acid sequence alignment of SEQ ID NO: 62 and 65 showing the differences in the 5' and 3' untranslated regions, wherein A=adenine, T=thymine, G=guanine, C=cytosine, N=any nucleic acid.

Figure 17A-F: Multiple nucleic acid sequence alignment of SEQ ID NO: 62 and 66 showing the differences in the 5' and 3' untranslated regions, wherein A=adenine, T=thymine, G=guanine, C=cytosine, N=any nucleic acid.

Figure 18: BLASTP amino acid sequence alignment of SEQ ID NO: 68 with chipmunk HP-20 precursor, gi:1170339 (SEQ ID NO: 79) showing 50% identity and 66% similarity over 153 amino acids of SEQ ID NO: 79.

Figure 19A: Multiple sequence alignment of adiponectin with sea urchin C1qDC proteins. A total of 5 closely-related C1qDC family members were identified in sea urchin. Their GenBank accession mumbers are AAK11302 [gi:12964750, Sp_C1qDC1 (SEQ ID NO: 80)], AAK11303 [gi:12964752, Sp_C1qDC2 (SEQ ID NO: 81)], AAG16425 [gi:10280597, Sp_C1qDC3 (SEQ ID NO: 82)], AAK11309 [gi:12964764, Sp_C1qDC4 (SEQ ID NO: 83)], AAK11305 [gi:12964756, Sp_C1qDC5 (SEQ ID NO: 84)]. Six of the 8 conserved residues in Figure 3 are conserved here as well. In the two other positions, a conservative replacement of phenylalanine (F) to tyrosine (Y) was seen in 2 and in 3 proteins, correspondingly.

Figure 19B: Multiple sequence alignment of adiponectin with Bacillus cereus C1qDC proteins. Three C1qDC proteins were identified, all with very low BLAST and Pfam scores. Their GenBank accession numbers are AAP09230 [gi:29895949, Bc_C1qDC1 (SEQ ID NO: 85)], AAP09231 [gi:29895950, Bc_C1qDC2 (SEQ ID NO: 86)], AAP09378 [gi:29896097, Bc_C1qDC3 (SEQ ID NO: 87)]. Bc_C1qDC1 and Bc_C1qDC2 are closely related, whereas Bc_C1qDC3 is much more divergent. Bc_C1qDC3 contains a collagen-repeat motif in the N-terminus preceding the C1q domain and is known as "collagen triple helix repeat protein" in GenBank. Five of the 8 conserved residues in Figure 3 are conserved here as well.

4. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to 14 novel C1q domain-containing polypeptides, herein denoted as CDCP.

Structural features of C1q domain-containing proteins

Figure 2 shows the schematic diagrams representing the domain structures and exon patterns of all the human C1q domain-containing proteins. Vertical lines indicate exon boundaries. Signal peptide domains are represented by the open bars and the number within the shaded bars represents the number of GXY repeats, wherein G represents Glycine and X and Y represent any amino acid. The C1q domain is represented by the black bars. The sizes of human

CDCP proteins vary from 193 amino acids (CBLN1) to 1228 amino acids (multimerin) with most of them ranging from 193 to 340 amino acids. Thirty of the 31 human proteins contain a single C-terminal C1q domain while C1qTNF4 contains 2 tandem C1q domains (Figure 2). All but one (C1qDC1) contains a leading signal peptide. To date, published reports have demonstrated secretion of twelve C1q-related proteins and the presence of leading signal peptides suggest that most, if not all, C1q-related proteins will be secreted. In the majority of human CDCP proteins, the signal peptide is followed by a collagen-like domain consisting of numerous repeats of the tripeptide GXY. The copy number of GXY repeats varies from 14 (C1qTNF1, 6, 8, and CRF2) to 153 (COL10A1). Long collagen regions are often broken into segments by imperfect GXY repeats, whereas short collagen regions maintain uninterrupted stretches of GXY repeats. COL8A1 and 2 and COL10A1 each contain 9 segments of GXY repeats and C1qA-C, AQL1-2, and otolin each contain 2 segments, whereas the rest of the C1q-related proteins contain one uninterrupted stretch of GXY repeats. The only exception is EMILIN2 whose 17 GXY repeats are interrupted 4 times by substituting glycine (G) with other residues (Doliana *et al., J. Biol. Chem.* 276:12003-12011 (2001) herein incorporated by reference in its entirety). The imperfect GXY repeats in C1qA-C may be important for the formation of the kinked collagen triple helix that assembles into the bouquet-like structure depicted in Figure 1 (Thiel and Reid, *FEBS Lett.* 250:78-84 (1989) herein incorporated by reference in its entirety). It is conceivable that disrupted collagen region may have more bending flexibility whereas perfect GXY repeats may form the rigid stalks of the collagen triple helix.

The highest sequence conservation among all C1q related proteins resides in the C1q domain. The percent identity numbers for pair-wise sequence alignment of different C1q domains can range from 20 to 40% for distantly related proteins and from 60 to 96% for closely related proteins (see Table 1).

Table 1: Similarity Matrix of Human C1q Domains

C1QTNF8	100	26	28.8	26.2	29.5	29.2	28	27.8	28	27.8	27.8	28	29.5	19.5	32.8
C1QTNF7		100	32.1	37.3	32.8	37.3	44.4	39.7	40.9	43.8	43.8	34.6	37.4	30.1	32.1
CRF2			100	34.9	88.8	36.9	33.3	33.3	34.9	32.6	32.6	83.2	79.2	23	33.6
Col8A1				100	35.7	72.8	60	39.2	42.1	43.8	43.8	37.2	35.7	28.7	33.3
CRF					100	36.9	35.7	36.9	34.9	35.6	35.6	88.8	84.8	23.3	35.2
Col8A2						100	64	38.4	40.5	49.2	47.7	38.5	39.2	28.8	33.3
Col10A1							100	41.6	46.8	47.3	47.3	35.2	36.4	26.2	31.8
Otolin-L								100	46.8	44.5	44.5	35.4	35.4	32.1	30.5
Adiponectin									100	55.8	55.8	37.2	38	29.5	29.8
AQL1										100	96.1	35.1	37.9	31.9	32.1
AQL2											100	35.1	37.9	31.1	32.1
Gliacolin												100	88.8	24	32.8
Gliacolin-L													100	26.5	35.9
Multimerin														100	31
EEG1L															100

A multiple sequence alignment of all 32 human C1q domains (two C1q domains from C1QTNF4) showed that there are 4 well-conserved regions separated by four less conserved regions. Most of the less conserved regions overlap with loop regions in the crystal structure (Figure 3). There are 15 highly conserved residues that are variable in four or less proteins (Figure 3, shaded residues). Among them, 8 residues are invariant for all human C1q domains (F115, F132, N138, F150, G156, Y158, F234, and G236, all are positions in adiponectin). In the crystal structure of adiponectin, the protein adopts a prototypic 10 β-strand jelly-roll with all 8 invariant residues found within the center of the structure (Figure 4). Among these 8 residues, all 5 aromatic residues are packed in the central hydrophobic core. Recently, two more C1q domain structures, from COL8A1 (Kvansakul *et al., Matrix Biol.* 22:145-152 (2003), herein incorporated by reference in its entirety) and COL10A1 (Bogin *et al., Structure (Camb.)* 10:165-173 (2002) herein incorporated by reference in its entirety), have become available. The locations of these 8 residues in these two structures are very similar to those in adiponectin (Figure 4). These highly conserved residues may play important roles in the formation or stabilization of the hydrophobic core of the C1q domain structure. However, if only the 10 β-strand jelly-roll folding model is considered, these residues are not irreplaceable. In the TNF family, which shares a highly similar folding topology, 3 out of the 5 aromatic residues are not conserved (Shapiro and Scherer, *Curr. Biol.* 8:335-338 (1998), herein incorporated by reference in its entirety). Thus, it is possible that these invariant C1q residues also play roles in maintaining a distinctive architecture or surface necessary in the function of all C1q proteins that clearly differs from the requirements of the related TNF family of proteins.

Subfamilies of C1q related proteins

Based on sequence homology, functional relatedness, and similarity in domain structure and intron-exon pattern (Figure 2), C1q-related proteins can be classified into multiple subsets. A subfamily grouping is described herein based on the phylogenetic tree of human CDCP proteins (Figure 5). Three major subfamilies could be readily identified, designated as the CDCP-A subfamily (the adiponectin/short collagen group), CDCP-B subfamily (the CBLN/gliacolin group), and the CDCP-C subfamily (the emilin/multimerin group).

Clustering of homologous genes in the adjacent chromosomal locations often indicates functional relatedness of the genes. Two such clusters are found among the 31 human CDCP encoding genes (Table 2). C1qA-C genes are clustered within a 25 kb region at chromosome

1p36.12 in the order of C1qA-C1qC-C1qB in the same orientation. The second gene cluster, involving AQL1 and AQL2, are separated by 420 kb on chromosome 13q12.12 with several potential intervening genes between them.

Table 2

Name	HUGO Symbol	Annotation	Chrom. Location	GenBank Accession	Human UniGene	Mouse UniGene	Hs_Mm id%
Adiponectin		adipocyte complement related protein of 30 kDa (ACRP30); adipoQ; adipose most abundant gene transcript 1 (APM1); gelatin-binding protein (GBP28)	3q27.3	NP_004788		Mm.3969	82
AQL1		adipoQ like 1	13q12.12	AAH40438	Hs.362854	Mm.59192	87
AQL2		adipoQ like 2	13q12.12	CAD57043		None	
C1QA	C1QA	complement component 1, q subcomponent, A chain	1p36.12	NP_057075	Hs.9641	Mm.370	70
C1QB	C1QB	complement component 1, q subcomponent, B chain	1p36.12	NP_000482	Hs.8986	Mm.2570	79
C1QC	C1QG	complement component 1, q subcomponent, C chain (gamma chain)	1p36.12	NP_758957	Hs.94953	Mm.3453	73
C1QDC1	C1QDC1	C1q domain containing 1 isoform L; EEG1L	12p11.21	NP_076414*	Hs.234355	Mm.3419	80
C1QTNF1	C1QTNF1	C1q and TNF related protein 1, G protein coupled receptor interacting protein (GIP), CTRP1, ZSIG37	17q25.3	NP_112230	Hs.201398	Mm.23845	77
C1QTNF2	C1QTNF2	C1q and TNF related protein 2, CTRP2, zacrp2	5q33.3	NP_114114	Hs.110062	Mm.24994	94
C1QTNF3	C1QTNF3	C1q and TNF related protein 3, collagenous repeat-containing sequence of 26-kDa (CORS26), CTRP3	5p13.2	NP_112207	Hs.171929	Mm.19310	95
C1QTNF4	C1QTNF4	C1q and TNF related protein 4, CTRP4, ZACRP4	11p11.2	NP_114115	Hs.119302	Mm.41630	95
C1QTNF5	C1QTNF5	C1q and TNF related protein 5, CTRP5	11q23.3	NP_056460	Hs.157211	Mm.137121	94
C1QTNF6	C1QTNF6	C1q and TNF related protein 6, CTRP6, ZACRP6	22q13.1	NP_114116*	Hs.22011	Mm.34776	67
C1QTNF7	C1QTNF7	C1q and TNF related protein 7, CTRP7, ZACRP7	4p15.33	NP_114117	Hs.153714	Mm.33391	96
C1QTNF8		C1q and TNF related protein 8; Similar to C1q and TNF related protein 6	16p13.3	XP_301604		None	
CBLN1	CBLN1	precerebellin 1	16q12.1	NP_004343	Hs.662	Mm.4880	99
CBLN2		ortholog of mouse precerebellin 2	18q22.3	AAH35789	Hs.7065	Mm.70775	94
CBLN3		ortholog of mouse precerebellin 3; similar to CBLN3	14q11.2	XP_292223		Mm.97163	93
CBLN4	CBLNL1	ortholog of mouse precerebellin 4; precerebellin-like 1 precursor	20q13.31	NP_542184	Hs.126141	Mm.40555	96
COL10A1	COL10A1	alpha-1 type X collagen	6q22.1	NP_000484	Hs.179729	Mm.4837	87
COL8A1	COL8A1	alpha-1 type VIII collagen	3q12.1	NP_001841	Hs.114599	Mm.86813	93
COL8A2	COL8A2	alpha-2 type VIII collagen	1p34.3	NP_005193	Hs.353001	Mm.29315	95
CRF1		C1q-related factor 1; C1q-related factor	17q21.31	NP_006679	Hs.134012	Mm.57154	99
CRF2		C1q-related factor 2; similar to C1q-related factor precursor	12q13.12	XP_290558	Hs.380386		96
EMILIN1	EMILIN1	elastin microfibril interfacer 1	2p23.3	NP_008977	Hs.63348	Mm.46229	86
EMILIN2	EMILIN2	elastin microfibril interfacer 2; Extracellular glycoprotein EMILIN-2 precursor	18p11.32	NP_114437	Hs.270143	Mm.23462	72
EMILIN3	EMILIN3	elastin microfibril interfacer 3; EMILIN-like protein EndoGlyx-1	10q23.2	NP_079032	Hs.127216	Mm.33798	65

Name	HUGO Symbol	Annotation	Chrom. Location	GenBank Accession	Human UniGene	Mouse UniGene	Hs_Mm Id%
Gliacolin1		ortholog of mouse Gliacolin; similar to Gliacolin	10p13	NP_872334		Mm.229322	99
Gliacolin2		C1q-domain containing protein; gliacolin-like	2q14.2	XP_092478	Hs.433493		94
Multimerin	MMRN	multimerin	4q22.1	NP_031377	Hs.268107	Mm.22904	65
Otolin		ortholog of salmon otolin; Similar to Otolin-1	3q26.1	XP_067228*			71

The prototypic C1qTNF proteins (C1qTNF-X) were identified by homology-based searches for TNF paralogs and do not constitute a discrete sub-family in the human complement. Since these names are approved by HUGO, they are used herein. C1qTNF members scatter within the first two subfamilies (Figure 5). Specifically, C1qTNF2, 5, and 7 are within the CDCP-A subfamily and C1qTNF1, 3, 4, 6, and 8 are found in the CDCP-B subfamily.

CDCP-A subfamily, the adiponectin/short collagen group

C1q subunits (C1qA, B, and C)

As mentioned above, the C1q domains of C1QA, B, and C form a heterotrimer. This trimerization is believed to mediate the formation of the triple helical collagen stalk (Kishore and Reid, 1999, *supra*). The heterotrimeric heads of C1q directly bind to the Fc region of aggregated IgG or IgM (Kishore and Reid, 1999, *supra*). Crosslinking experiments suggest that all three subunits are involved in binding (Wines and Easterbrook-Smith, *Mol. Immunol.* 27:221-226 (1990), herein incorporated by reference in its entirety). Individual recombinant C1q domains of C1QA, B, and C, as monomers, can bind preferentially to either IgG (C1QB), or IgM (C1QC), or both (C1QA) *in vitro* (Kishore *et al., J. Immunol.* 166:559-565 (2001); Kishore *et al., J. Immunol.* 171:812-820 (2003), both of which are herein incorporated by reference in their entirety). Recombinant C1q domains of C1QA and B were also shown to inhibit C1q-mediated hemolysis of IgG- and IgM-sensitized sheep erythrocytes (Kishore *et al.*, 2001, 2003, *supra*). These results suggest that each C1q domain seems to keep relative structural and functional independence. When associated together, each of them contributes to the functional multivalency and flexibility of the heterotrimer.

Adiponectin

Mouse and human adiponectins were identified independently by four laboratories and was named Acrp30 (Scherer *et al., J. Biol. Chem.* 270:26746-26749 (1995), AdipoQ (Hu *et al.,*

J. Biol. Chem. 271:10697-10703 (1996), APM1 (Maeda *et al., Biochem. Biophys. Res. Commun.* 221:286-289 (1996), and GBP28 (Nakano *et al., J. Biochem. (Tokyo)* 120:803-812 (1996)), respectively (these and all references are herein incorporated by reference in their entirety). Adiponectin has been shown to increase insulin sensitivity as well as regulate lipid and glucose metabolism, and has anti-inflammatory and anti-atherogenic properties (for review see Berg *et al., Trends Endocrinol. Metab.* 13:84-89 (2002); Tsao *et al., Eur. J. Pharmacol.* 440:213-221 (2002); Stefan and Stumvoll, *Horm. Metab. Res.* 34:469-474 (2002); Diez and Iglesias, *Eur. J. Endocrinol.* 148:293-300 (2003); Pajvani *et al., J. Biol. Chem.* 278:9073-9085 (2003), all of which are herein incorporated by reference in their entirety). It is the most abundant protein expressed specifically in adipose tissue. The concentrations of adiponectin in human plasma range from 5 to 30 mg/ml, accounting for ~ 0.01 to 0.05% of total plasma protein (Diez and Iglesias, 2003, *supra*; Scherer *et al.*, 1995, *supra*). This concentration is unusually high for a hormone (3 orders of magnitude higher than most hormones). Recently, a proteolytic product containing essentially only the C1q domain of adiponectin was shown to increase free fatty acid oxidation in muscle and cause weight loss in mice (Fruebis *et al., Proc. Natl. Acad. Sci. USA* 98:2005-2010 (2001), herein incorporated by reference in its entirety). Remarkably, this proteolytic product is much more potent than the intact adiponectin in causing these effects, suggesting that processed adiponectin is the bioactive hormone.

AQL1 and AQL2

AQL1 and AQL2 are almost identical, with 99% identity at nucleotide level in the coding region and 98% identity at amino acid level. Only 7 out of the 333 residues are different, with 6 of them located in the C1q domain. The upstream 9kb sequences are also well conserved for the two genes, with ~80-90% identity. They are located closely on chromosome 13q12.12, separated by only 400 kb, with the same intron-exon pattern (Figure 2). Interestingly, only one copy of such gene was found in mouse and rat. The mouse and rat orthologs are closer to AQL1 than to AQL2. It appears that AQL2 is derived from AQL1 during a very recent duplication event. Expression profiling data indicate that both genes are expressed in similar tissues (skeletal muscle, heart, and adipose tissue).

Although named after adiponectin, AQL1 and AQL2 are not exclusively expressed in adipose tissue. They also have a much longer collagen-like region (56 GXYs in two stretches) compared to adiponectin (22 GXYs in one stretch). In addition, the sequence homology level between adiponectin and AQLs is not very high (56% identity for the C1q domain).

The present invention relates to four CDCP polypeptides that are homologous to adiponectin: SEQ ID NOs: 4, 7, 10, and 19. The first adiponectin-like CDCP polypeptide of SEQ ID NO: 4 is an approximately 288 amino acid protein with a predicted molecular mass of approximately 32 kD unglycosylated. The initial methionine starts at position 18 of SEQ ID NO: 3 and the putative stop codon begins at position 882 of SEQ ID NO: 3. Protein database searches with the BLASTP algorithm (Altschul *et al., J. Mol. Evol.* 36:290-300 (1993) and Altschul *et al., J. Mol. Biol.* 21:403-410 (1990), both of which are herein incorporated by reference in their entirety) indicate that SEQ ID NO: 4 shares 86% identity with similar-to-adiponectin precursor (ACRP30), gi:29738938 (SEQ ID NO:70) over 333 amino acids of SEQ ID NO: 70.

Using the pfam software program (Sonnhammer *et al., Nucl. Acids Res.* 26:320-322 (1998), herein incorporated by reference in its entirety), the CDCP polypeptide of SEQ ID NO: 4 revealed its structural homology to C1q and collagen domains (see Table 3). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 3

e-value	Score	Model	Description	Amino acid position
8.8e-12	48.3	Collagen	Collagen triple helix repeat (20 copies)	24-82
2.5e-10	43.0	Collagen	Collagen triple helix repeat (20 copies)	13-172
3.2e-38	140.4	C1q	C1q domain	173-284

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al., J. Comp. Biol.* 6:219-235 (1999), herein incorporated by reference in its entirety), the CDCP polypeptide of SEQ ID NO: 4 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 4). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 4

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	7-194
IPB000885B	Fibrillar collagen C-terminal domain	9-197
IPB001073B	Complement C1q protein	28-286
PRO1525F	EDG-5 sphingosine 1-phosphate receptor signature VI	32-42
IPB00817A	Prion protein	122-164
PRO0007A	Complement C1q domain signature I	169-195
PRO0007B	Complement C1q domain signatuare II	196-215
PRO0007C	Complement C1q domain signature III	240-261
PRO0007D	Complement C1q domain signature IV	275-285

The second adiponectin-like CDCP polypeptide of the invention (SEQ ID NO: 7) is an approximately 300 amino acid protein with a predicted molecular mass of approximately 34 kD unglycosylated. The initial methionine starts at position 18 or SEQ ID NO: 6 and the putative stop codon begins at position 918 of SEQ ID NO: 6. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra* and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 7 shares 89% identity and 90% similarity with similar-to-ACRP30 (SEQ ID NO:70) over 302 amino acids of SEQ ID NO: 70.

Using the pfam software program (Sonnhammer *et al.*, 1998, *supra*) the CDCP polypeptide of SEQ ID NO: 7 revealed its structural homology to C1q and collagen domains (see Table 5). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 5

e-value	Score	Model	Description	Amino acid position
8.8e-12	48.3	Collagen	Collagen triple helix repeat (20 copies)	24-82
2.9e-10	42.8	Collagen	Collagen triple helix repeat (20 copies)	95-154
9.8e-08	33.6	Collagen	Collagen triple helix repeat (20 copies)	155-191
3.7e-08	35.8	C1q	C1q domain	227-275

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 7 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 6). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 6

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	7-212
IPB000885B	Fibrillar collagen C-terminal domain	9-221
IPB00173A	Complement C1q protein	28-277
PR01525F	EDG-5 sphingosine 1-phosphate receptor signature VI	32-42
IPB000817A	Prion protein	122-164
PR00007C	Complement C1q domain signature III	258-279

The third adiponectin-like CDCP polypeptide of the invention (SEQ ID NO: 10) is an approximately 333 amino acid protein with a predicted molecular mass of approximately 35 kD unglycosylated. The initial methionine starts at position 25 or SEQ ID NO: 9 and the putative stop codon begins at position 1024 of SEQ ID NO: 9. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra* and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 10 shares 99% identity with similar-to-ACRP30 (SEQ ID NO: 70) over 333 amino acids of SEQ ID NO: 70.

Using the pfam software program (Sonnhammer *et al.*, 1998, *supra*) the CDCP polypeptide of SEQ ID NO: 10 revealed its structural homology to C1q and collagen domains (see Table 7). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 7

e-value	Score	Model	Description	Amino acid position
8.8e-12	48.3	Collagen	Collagen triple helix repeat (20 copies)	24-82
2.9e-10	42.8	Collagen	Collagen triple helix repeat (20 copies)	95-154
6.6e-08	34.2	Collagen	Collagen triple helix repeat (20 copies)	155-191
9.2e-42	152.2	C1q	C1q domain	203-329

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 10 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 8). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 8

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	7-212
IPB000885B	Fibrillar collagen C-terminal domain	9-221
IPB00173B	Complement C1q protein	28-331
PR01525F	EDG-5 sphingosine 1-phosphate receptor signature VI	32-42
IPB000817A	Prion protein	122-164
PR00007A	Complement C1q domain signature I	214-240
PR00007B	Complement C1q domain signature II	241-260
PR00007C	Complement C1q domain signature III	285-306
PR00007D	Complement C1q domain signature IV	320-330

The fourth adiponectin-like CDCP polypeptide of the invention (SEQ ID NO: 19) is an approximately 306 amino acid protein with a predicted molecular mass of approximately 34 kD unglycosylated. The initial methionine starts at position 25 or SEQ ID NO: 18 and the putative stop codon begins at position 943 of SEQ ID NO: 18. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra* and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 10 shares 91% identity with similar-to-ACRP30 (SEQ ID NO: 70) over 333 amino acids of SEQ ID NO: 70.

Using the pfam software program (Sonnhammer *et al.*, 1998, *supra*) the CDCP polypeptide of SEQ ID NO: 19 revealed its structural homology to C1q and collagen domains (see Table 9). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 9

e-value	Score	Model	Description	Amin acid position
8.8e-12	48.3	Collagen	Collagen triple helix repeat (20 copies)	24-82
2.9e-10	42.8	Collagen	Collagen triple helix repeat (20 copies)	95-154
6.6e-08	34.2	Collagen	Collagen triple helix repeat (20 copies)	155-191
1.6e-15	63.5	C1q	C1q domain	227-302

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 19 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 10). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 10

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	7-212
IPB000885B	Fibrillar collagen C-terminal domain	9-221
IPB00173A	Complement C1q protein	28-304
PR01525F	EDG-5 sphingosine 1-phosphate receptor signature VI	32-42
IPB000817A	Prion protein	122-164
PR00007C	Complement C1q domain signature III	258-279
PR00007D	Complement C1q domain signature IV	293-303

All four adiponectin-like CDCP polypeptides are prediced to contain an approximately nineteen (19) residue signal peptide is encoded from approximately amino acids 1 to 19 of SEQ ID NO: 4, 7, 10 and 19. The extracellular portions are useful on their own. The signal peptide region was predicted using the Neural Network SignalP V1.1 program (Nielsen *et al., Int. J. Neural Syst.* 8:581-599 (1997) herein incorporated by reference in its entirety). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program. SEQ ID NO: 5 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 4. SEQ ID NO: 8 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 7. SEQ ID NO: 11 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 10. SEQ ID NO: 20 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 19.

Figure 6 shows a Clustal-W multiple amino acid sequence alignment of SEQ ID NOs: 4, 7, 10, and 19 with human similar-to-ACRP30, gi:29738938 (SEQ ID NO: 70), wherein identical residues are represented by an asterisk (*), conservative substitutions are represented by a colon (:), and semi-conservative substitutions are represented by a period (.).

The adiponectin-like CDCP polypeptides of the invention are expected to have activity similar to adiponectin. Therefore, they are expected to be useful in the treatment, amelioration,

and/or diagnosis of diseases and disorders relating to lipid metabolism and/or glucose metabolism, cardiovascular diseases, diabetes, stroke, obesity, and the like.

Short chain collagens

The short chain collagens COL8A1, COL8A2, and COL10A1 share many similarities including intron-exon pattern, domain structure, and lengths of their collagen-like regions (151, 152, and 153 GXY repeats, respectively). They all contain 9 stretches of GXY repeats with a similar fragmentation pattern and exist as homotrimers in tissue (Greenhill *et al., Matrix Biol.* 19:19-28 (2000); Bogin et al., 2002, *supra*, herein incorporated by reference in their entirety). They can also form higher order polygonal lattices (Sawada *et al., J. Cell Biol.* 110:219-227 (1990); Kwan *et al., J. Cell Biol.* 114:597-604 (1991), herein incorporated by reference in their entirety). Crystal structures of both COL8A1 and COL10A1 reveal the presence of 3 aromatic stripes on the surface of the trimer, which may play important roles in higher order assembly of these molecules (Kvansakul *et al., Martix Biol.* 22:145-152 (2003); Bogin *et al.*, 2002, *supra*; herein incorporated by reference in their entirety). A buried cluster of calcium ions is found in the structure of the COL10A1 trimer, which may contribute to the stability of the assembly (Bogin *et al.*, 2002, *supra*). Calcium ions are not present in structures of adiponectin and COL8A1. Mutations in COL8A2 have been shown to cause two types of corneal endothelial dystrophy (Biswas *et al., Hum. Mol. Genet.* 10:2415-2423 (2001) herein incorporated by reference in its entirety). Mutations in COL10A1, of which most are in C1q domain, have been demonstrated to cause Schmid metaphyseal chondrodysplasia (summarized in Bogin *et al.*, 2002, *supra*).

The present invention also relates to one short chain collagen, SEQ ID NO: 24, which is an approximately 744 amino acid protein with a predicted molecular mass of approximately 83 kD unglycosylated. The initial methionine starts at position 235 of SEQ ID NO: 23 and the putative stop codon begins at position 2467 of SEQ ID NO: 23. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra* and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 24 shares 99% identity with to human α1 type VIII collagen precursor, gi:17738302 (SEQ ID NO: 71) over 744 amino acids of SEQ ID NO: 71 (see Figure 7).

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 24 revealed its structural homology to C1q and collagen domains

(see Table 11). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 11

e-value	Score	Model	Description	Amino acid position
3.3e-08	35.3	Collagen	Collagen triple helix repeat (20 copies)	158-206
1.8e-05	25.3	Collagen	Collagen triple helix repeat (20 copies)	208-245
4.3e-06	27.6	Collagen	Collagen triple helix repeat (20 copies)	272-314
3e-09	39.1	Collagen	Collagen triple helix repeat (20 copies)	357-416
1.3e-10	44.1	Collagen	Collagen triple helix repeat (20 copies)	423-473
3e-10	42.7	Collagen	Collagen triple helix repeat (20 copies)	474-531
9.3e-05	22.7	Collagen	Collagen triple helix repeat (20 copies)	533-571
3.4e-74	259.9	C1q	C1q domain	617-741

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 24 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 12). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 12

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	87-602
IPB000885B	Fibrillar collagen C-terminal domain	89-602
IPB001073B	Complement C1q protein	111-743
IPB00817A	Prion protein	138-573
IPB001359H	Synapsin	447-571
PR00049D	Wilm's tumour protein signature IV	557-580
PR00007A	Complement C1q domain signature I	626-652
PR00007B	Complement C1q domain signature II	653-672
PR00007C	Complement C1q domain signature III	698-719
PR00007D	Complement C1q domain signature IV	732-742

A predicted approximately 27 residue signal peptide is encoded from approximately residue 1 to residue 27 of SEQ ID NO: 24. The extracellular portion is useful on its own. The signal peptide region was predicted using the Neural Network SignalP V1.1 program (Nielsen *et al.*, 1997, *supra*). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program. SEQ ID NO: 25 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 24.

The short chain collagen-like CDCP polypeptide of the invention is expected to have activity similar to other short chain collagens. Therefore, it is expected that SEQ ID NO: 24 will be useful as a therapeutic and/or diagnostic for disorders and diseases associated with extracellular matrix abnormalities, including but not limited to disorders of the cornea, chondrodysplasias, and other collagen-related disorders.

Otolin

Otolin, also known as inner ear specific-collagen and saccular collagen, was first identified in bluegill sunfish by differential screening for saccule-specific cDNAs (Davis *et al.*, *Science* 267:1031-1034 (1995) herein incorporated by reference in its entirety). It was later found as a major structural protein in chum salmon otolith, a calcified organ in the inner ear that functions in the hearing and balancing systems, and thus was named otolin (Murayama *et al., Eur. J. Biochem.* 269:688-696 (2002) herein incorporated by reference in its entirety). Otolin is specifically expressed in the sacculus, synthesized in the transitional epithelium and transferred to the otolith and otolithic membrane.

The best human homolog of otolin is a predicted gene, similar to otolin-1 (GenBank accession XP_067228) that aligns to the majority, but not full length, of salmon otolin. Based on sequences of salmon otolin, two mouse inner ear ESTs, and murine and human genomic sequences, Applicants re-edited human otolin to full length and is represented as SEQ ID NO: 91 in the sequence listing. Furthermore, the predicted murine and rat otolin (from GenBank submissions) were re-edited and are represented as SEQ ID NOs: 92 and 93. In addition, Applicants predicted the Fugu fish otolin gene based on Fugu genomic sequence (SEQ ID NO: 88). The original bluegill sunfish otolin, however, did not align well with the rest. Closer examination revealed that when 3 single nucleotides were added at 3 positions into the original cDNA sequence, the resulting translation product would align very well with other otolins (Figure 8). This is probably due to the poor sequencing quality in regions of the original cDNA, rather than representing a real difference between bluegill sunfish and other species.

The present invention relates to one otolin-like CDCP polypeptide (SEQ ID NO: 27). SEQ ID NO: 27 is an approximately 477 amino acid protein with a predicted molecular mass of approximately 52 kD unglycosylated. The initial methionine starts at position 9 of SEQ ID NO: 26 and the putative stop codon begins at position 1440 of SEQ ID NO: 26. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 27 shares 94% identity with human similar to otolin-1, gi:22041493 (SEQ ID NO: 78) over 459 amino acids of SEQ ID NO: 78 (see Figure 9).

Using the Pfam software program (Sonnhammer et al., 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 27 revealed its structural homology to C1q domains (see Table 13). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 13

e-value	Score	Model	Description	Amino acid position
2.7e-05	22.2	Collagen	Collagen triple helix repeat (20 copies)	109-146
9.4e-10	38.9	Collagen	Collagen triple helix repeat (20 copies)	149-197
1.4e-05	23.2	Collagen	Collagen triple helix repeat (20 copies)	209-242
2.2e-10	41.2	Collagen	Collagen triple helix repeat (20 copies)	245-304
3.4e-04	18.0	Collagen	Collagen triple helix repeat (20 copies)	305-335
1.9e-34	124.6	C1q	C1q domain	344-467

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 27 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 14). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 14

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	88-359
IPB000885B	Fibrillar collagen C-terminal domain	90-359
IPB001073B	Complement C1q protein	94-469
IPB000817A	Prion protein	141-189
PR00007A	Complement C1q domain signature I	353-379
PR00007B	Complement C1q domain signature II	380-399
PR00007C	Complement C1q domain signature III	425-446
PR00007D	Complement C1q domain signature IV	458-468

A predicted approximately 18 residue signal peptide is encoded from approximately residue 1 to residue 18 of SEQ ID NO: 27. The extracellular portion is useful on its own. The signal peptide region was predicted using the Neural Network SignalP V1.1 program (Nielsen *et al.*, 1997, *supra*). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program. SEQ ID NO: 28 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 27.

The otolin-like CDCP polypeptide of the invention is expected to have properties and activities similar to that of other members of the otolin family. Therefore, it is expected that SEQ ID NO: 27 will be useful in treating disorders and diseases associated with hearing and balance and abnormalities of the cochlear structures.

CDCP-B subfamily, the CBLN/gliacolin group

Precerebellins

There are four members in the precerebellin subfamily and they share highly homologous sequences and similar intron-exon patterns and domain structures (Figure 2). Three of these or their mouse orthologs (CBLN1-3) have been described in the literature (Urade *et al.*, 1991, *supra*; Wada

and Ohtani, *Brain Res. Mol. Brain Res.* 9:71-77 (1991); Pang *et al., J. Neurosci.* 20:6333-6339 (2000), herein incorporated by reference in their entirety). There are many human and mouse ESTs supporting CBLN4 as an actively transcribed gene. All 4 genes are expressed in neuronal tissues but maintain distinctive expression patterns. For example, CBLN1 is expressed mainly in the adult cerebellum, whereas CBLN2 is expressed in extracerebellar brain areas and in fetal brain (Urade *et al.*, 1991, *supra*; Wada and Ohtani, 1991, *supra*). CBLN3 shows a very similar temporal and spatial expression pattern as that of CBLN1, and was demonstrated to interact with CBLN1 in the yeast two-hybrid system (Pang *et al.*, 2000, *supra*). Therefore, CBLN1 and 3 may form heterotrimers in vivo. With the triple helical collagen regions absent, the presumed CBLN trimers are probably less stable than those of other C1q-related proteins with collagen-like regions. The 16-mer cerebellin peptide partially overlaps with the N-terminal end of the C1q domain, including 2 of the 15 highly conserved residues. Therefore, processing of the cerebellin peptide may significantly affect the stability of the trimeric structure.

Gliacolins and CRFs

Members in this subfamily have the highest sequence conservation among all C1q-related proteins, with the same intron-exon pattern and domain structure. All of them contain a short stretch of GXY repeats in the collagen-like region. Gliacolin was identified in a yeast two-hybrid screen to interact with a chaperone protein that is known to bind collagen-like regions (Koide *et al., J. Biol. Chem.* 275:27957-27963 (2000) herein incorporated by reference in its entirety). The CRF gene was isolated from a cosmid library in a screen to identify genes involved in cellular senescence (Berube *et al., Brain Res. Mol. Brain Res.* 63:233-240 (1999) herein incorporated by reference in its entirety). It was shown to be expressed mainly in areas of the nervous system involved in motor function.

The present invention also relates to seven (7) CDCP polypeptides that are part of the gliacolin/CRF subfamily. The first, SEQ ID NO: 32, is an approximately 338 amino acid protein with a predicted molecular mass of approximately 38 kD unglycosylated. The initial methionine starts at position 199 of SEQ ID NO: 31 and the putative stop codon begins at position 1213 of SEQ ID NO: 31. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 32 shares 94% identity with murine gliacolin, gi:23680960 (SEQ ID NO: 72) over 255 amino acids of SEQ ID NO: 72 and

70% identity and 78% similarity with human C1q-related factor, gi:5729785 (SEQ ID NO: 73) over 262 amino acids of SEQ ID NO: 73.

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 32 revealed its structural homology to C1q and collagen domains (see Table 15). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 15

e-value	Score	Model	Description	Amino acid position
6.4e-08	34.2	Collagen	Collagen triple helix repeat (20 copies)	144-192
2.7e-31	117.4	C1q	C1q domain	211-335

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 32 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 16). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 16

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	116-222
IBP000885B	Fibrillar collagen C-terminal domain	115-222
IPB001073A	Complement C1q protein	137-337
PB000817A	Prion protein	145-193
PR00007A	Complement C1q domain signature I	219-245
PR00007B	Complement C1q domain signature II	246-265
PR00007C	Complement C1q domain signature III	294-315
PR00007D	Complement C1q domain signature IV	326-336

The second gliacolin/CRF-like CDCP polypeptide of the invention (SEQ ID NO: 34) is an approximately 244 amino acid protein with a predicted molecular mass of approximately 27 kD unglycosylated. The initial methionine starts at position 161 of SEQ ID NO: 33 and the putative stop codon begins at position 893 of SEQ ID NO: 33. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 34 shares 94% identity with murine gliacolin (SEQ ID NO: 72) over 255 amino acids of SEQ ID NO: 72 and 70% identity and 78% similarity with human C1q-related factor (SEQ ID NO: 73) over 262 amino acids of SEQ ID NO: 73.

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 34 revealed its structural homology to C1q and collagen domains (see Table 17). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 17

e-value	Score	Model	Description	Amino acid position
6.4e-08	34.2	Collagen	Collagen triple helix repeat (20 copies)	50-98
2.7e-31	117.4	C1q	C1q domain	117-241

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 34 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 18). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 18

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	21-128
IBP000885B	Fibrillar collagen C-terminal domain	22-128
IPB001073B	Complement C1q protein	43-243
PB000817A	Prion protein	51-99
PR00007A	Complement C1q domain signature I	125-151
PR00007B	Complement C1q domain signature II	152-171
PR00007C	Complement C1q domain signature III	200-221
PR00007D	Complement C1q domain signature IV	232-242

A predicted approximately 19 residue signal peptide is encoded from approximately residue 1 to residue 19 of SEQ ID NO: 34. The extracellular portion is useful on is own. The signal peptide region was predicted using the Neural Network SignalP V1.1 program (Nielsen *et al.*, 1997, *supra*). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program. SEQ ID NO: 35 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 34.

The third CDCP polypeptide of of the invention of the gliacolin subfamily (SEQ ID NO: 38) is an approximately 293 amino acid protein with a predicted molecular mass of approximately 33 kD unglycosylated. The initial methionine starts at position 683 of SEQ ID NO: 37 and the putative stop codon begins at position 1562 of SEQ ID NO: 37. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 38 shares 64% identity and 71% similarity with murine gliacolin (SEQ ID NO: 72) over 179 amino acids of SEQ ID NO: 72 and 62% identity and 69% similarity with human C1q-related factor (SEQ ID NO: 73) over 196 amino acids of SEQ ID NO: 73.

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 38 revealed its structural homology to C1q and collagen domains

(see Table 19). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 19

e-value	Score	Model	Description	Amino acid position
3.6e-05	24.2	Collagen	Collagen triple helix repeat (20 copies)	53-96
2.8e-13	55.1	C1q	C1q domain	111-178

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 38 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 20). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 20

Accession number	Name	Amino acid position
IBP000885A	Fibrillar collagen C-terminal domain	29-124
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	30-127
IPB001073B	Complement C1q protein	45-160
IPB002896E	Herpesvirus glycoprotein D	65-102
IPB001359H	Synapsin	69-119
PB000817A	Prion protein	52-101
PR00007A	Complement C1q domain signature I	125-151
PR00007B	Complement C1q domain signature II	152-171

The fourth gliacolin-like CDCP polypeptide of SEQ ID NO: 41 is an approximately 238 amino acid protein with a predicted molecular mass of approximately 27 kD unglycosylated. The initial methionine starts at position 683 of SEQ ID NO: 40 and the putative stop codon begins at position 1397 of SEQ ID NO: 40. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: shares 70% identity and 77% similarity with murine gliacolin (SEQ ID NO: 72) over 238 amino acids of SEQ ID NO: 72 and 69% identity and 74% similarity with human C1q-related factor (SEQ ID NO: 73) over 258 amino acids of SEQ ID NO: 73.

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 41 revealed its structural homology to C1q and collagen domains (see Table 21). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 21

e-value	Score	Model	Description	Amino acid position
3.6e-05	24.2	Collagen	Collagen triple helix repeat (20 copies)	53-96
6.5e-29	109.5	C1q	C1q domain	111-235

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 41 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 22). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 22

Accession number	Name	Amino acid position
IBP000885A	Fibrillar collagen C-terminal domain	29-124
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	30-127
IPB001073B	Complement C1q protein	45-237
PB000817A	Prion protein	52-101
IPB002896E	Herpesvirus glycoprotein D	65-102
IPB001359H	Synapsin	69-119
PR00007A	Complement C1q domain signature I	119-145
PR00007B	Complement C1q domain signature II	146-165
PR00007C	Complement C1q domain signature III	194-215
PR00007D	Complement C1q domain signature IV	226-236

A predicted approximately 15 residue signal peptide is encoded from approximately amino acid 1 to 15 of SEQ ID NO: 38 or 41. The extracellular portion is useful on its own. The signal peptide region was predicted using the Neural Network SignalP V1.1 program (Nielsen *et al.*, 1997, *supra*). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program. SEQ ID NO: 39 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 38. SEQ ID NO: 42 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 41.

Figure 10 shows a Clustal-W multiple amino acid sequence alignment of SEQ ID NOs: 32, 34, 38, and 41 with murine gliacolin (gi:23680960) (SEQ ID NO: 72), wherein identical residues are represented by an asterisk (*), conservative substitutions are represented by a colon (:), and semi-conservative substitutions are represented by a period (.). Gaps are represented as dashes.

Figure 11 shows a Clustal-W multiple amino acid sequence alignment of SEQ ID NOs: 32, 34, 38, and 41 with human C1q-related factor (SEQ ID NO: 73), wherein identical residues are represented by an asterisk (*), conservative substitutions are represented by a colon (:), and semi-conservative substitutions are represented by a period (.). Gaps are represented as dashes

The fifth gliacolin/CRF-like CDCP polypeptide of the invention (SEQ ID NO: 46) is an approximately 800 amino acid protein with a predicted molecular mass of approximately 90 kD unglycosylated. The initial methionine starts at position 511 of SEQ ID NO: 45 and the putative stop codon begins at position 2911 of SEQ ID NO: 45. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that

SEQ ID NO: 46 shares 85% identity and 86% similarity with human C1q domain-containing 1 isoform L (EEG1L), gi:23503235 (SEQ ID NO: 74), over 952 amino acids of SEQ ID NO: 74.

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP gliacolin-like polypeptide of SEQ ID NO: 46 revealed its structural homology to C1q and collagen domains (see Table 23). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 23

e-value	Score	Model	Description	Amino acid position
1.6e-26	101.5	C1q	C1q domain	672-797

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 46 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 24). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 24

Accession number	Name	Amino acid position
IPB002360C	Involucrin	151-192
PR00007A	Complement C1q domain signature I	683-709
IPB001073B	Complement C1q protein	690-799
PR00007B	Complement C1q domain signature II	710-729
PR00007C	Complement C1q domain signature III	757-778
PR00007D	Complement C1q domain signature IV	788-798

The sixth member of the gliacolin subfamily is the CDCP polypeptide of SEQ ID NO: 48 which is an approximately 710 amino acid protein with a predicted molecular mass of approximately 80 kD unglycosylated. The initial methionine starts at position 511 of SEQ ID NO: 47 and the putative stop codon begins at position 2641 of SEQ ID NO: 47. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 48 shares 95% identity with human EEG1L (SEQ ID NO: 74) over 892 amino acids of SEQ ID NO: 74.

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), CDCP polypeptide of SEQ ID NO: 48 revealed its structural homology to C1q and collagen domains (see Table 25). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 25

e-value	Score	Model	Description	Amino acid position
2e-20	81.3	C1q	C1q domain	582-707

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 48 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 26). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 26

Accession number	Name	Amino acid position
IPB002360C	Involucrin	151-192
IPB001073B	Complement C1q protein	600-709
PR00007B	Complement C1q domain signature II	620-639
PR00007C	Complement C1q domain signature III	667-688
PR00007D	Complement C1q domain signature IV	698-708

The seventh gliacolin-like CDCP polypeptide of SEQ ID NO: 51 is an approximately 1045 amino acid protein with a predicted molecular mass of approximately 115 kD unglycosylated. The initial methionine starts at position 241 of SEQ ID NO: 50 and the putative stop codon begins at position 3376 of SEQ ID NO: 50. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 51 shares 90% identity and 91% similarity to EEG1L (SEQ ID NO: 74) over 1059 amino acids of SEQ ID NO: 74.

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 51 revealed its structural homology to C1q domains (see Table 27). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 27

e-value	Score	Model	Description	Amino acid position
1.7e027	101.5	C1q	C1q domain	917-1042

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 51 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 28). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 28

Accession number	Name	Amino acid position
IPB002360C	Involucrin	403-444
PR00007A	Complement C1q domain signature I	928-954
IPB001073B	Complement C1q protein	935-1044
PR00007B	Complement C1q domain signature II	955-974
PR00007C	Complement C1q domain signature III	1002-1023
PR00007D	Complement C1q domain signature IV	1033-1043

Figure 12 shows a Clustal-W multiple amino acid sequence alignment of SEQ ID NOs: 46, 48, and 51 with human C1q domain-containing 1 isoform L (EEG1L), gi:23503235 (SEQ ID NO: 74), wherein identical residues are represented by an asterisk (*), conservative substitutions are represented by a colon (:), and semi-conservative substitutions are represented by a period (.). Gaps are represented as dashes.

The gliacolin-like CDCP polypeptides of the invention are expected to have the same properties and activities as gliacolin. Therefore, it is expected that the gliacolin-like polypeptides of the invention will be useful as therapeutics and/or diagnostics in disorders and diseases involving cellular senescence and neurological disorders, including, but not limited to, disorders in motor function.

CDCP-C subfamily, the EMILIN/multimerin group

EMILINs and multimerin

EMILIN1, 2, 3, and multimerin are large proteins of ~1000 aa or longer. They share the following domain organization: an N-terminal cysteine rich EMI domain followed by an extended region containing sequence elements with high potential of forming coiled-coil structure, and a C-terminal C1q domain. In addition, EMILIN1 and EMILIN2 contain a short collagen-like region adjacent to the C1q domain, whereas EMILIN3 and multimerin do not. EMILIN1 is an extracellular matrix component associated with elastic fibers (Doliana *et al.*, 1999, *supra*). It is highly expressed in blood vessels, skin, heart, and lung. It was reported recently that cell adhesion to EMILIN1 is mediated by its C1q domain (Spessotto *et al., J. Biol. Chem.* 278:6160-6167 (2002) herein incorporated by reference in its entirety). EMILIN2 was identified by a yeast two-hybrid screen using the C1q domain of EMILIN1 as bait (Doliana *et al.*, 2001, *supra*); however, EMILIN 1 and 2 are not co-expressed. EMILIN2 is mainly expressed in the cochlear basilar membrane and may be involved in auditory function (Amma *et al., Mol. Cell Neurosci.* 23:460-472 (2003) herein incorporated by reference in its entirety). The EMILIN3 gene codes for at least 2 of the 4 subunits in EndoGlyx-1, a cell surface glycoprotein complex found exclusively on blood vessel endothelium (Christian *et al., J. Biol. Chem.* 276:48588-48595 (2001) herein incorporated by reference in its entirety). Multimerin is a massive homomultimeric protein associated with coagulation protein factor V found in platelet α-granules and in vascular endothelium (Hayward *et al., J. Biol. Chem.* 270:18246-18251

(1995); Hayward *et al., J. Biol. Chem.* 270:19217-19224 (1995), herein incorporated by reference in their entirety).

In addition, the present invention relates to one EMILIN-like CDCP polypeptide (SEQ ID NO: 55). SEQ ID NO: 55 is an approximately 513 amino acid protein with a predicted molecular mass of approximately 57 kD unglycosylated. The initial methionine starts at position 1 of SEQ ID NO: 54 and the putative stop codon begins at position 1540 of SEQ ID NO: 54. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, supra) indicate that SEQ ID NO: 55 shares 98% identity with human EMILIN-2 precursor, gi:14042988 (SEQ ID NO:77) over 267 amino acids of SEQ ID NO: 77 (see Figure 13).

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 55 revealed its structural homology to C1q domains (see Table 29). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 29

e-value	Score	Model	Description	Amino acid position
1.5e-08	37.3	C1q	C1q domain	367-412

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 55 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 30). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 30

Accession number	Name	Amino acid position
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	278-363
IPB000885B	Fibrillar collagen C-terminal domain	280-363
IPB001073B	Complement C1q protein	290-418
PR00007A	Complement C1q domain signature I	377-403

The EMILIN-like CDCP polypeptide of the invention is expected to possess the same properties and activities as EMILIN polypeptides. Therefore, SEQ ID NO: 55 is expected to be useful in treating conditions relating to extracellular matrix disorders, auditory disorders, cardiovascular diseases, thromboses, and vascular disorders associated with platelets and coagulation.

Other CDCP proteins

C1qTNFs

There are currently 8 human C1qTNFs. C1qTNF1 was identified in a yeast two-hybrid screen using an intracellular loop region from a G-protein coupled receptor as bait, and therefore was also named GIP for "GPCR interacting protein" (Innamorati *et al., Regul. Pept.* 109:173-179 (2002) herein incorporated by reference in its entirety). It is predominantly expressed in heart whereas the GPCR that interacts with it is mainly expressed in kidney. C1qTNF1 has potent anti-thrombic activities and is currently in clinical evaluation (Zymogenetics product candidate described on the Zymogenetics website, Seattle, WA). C1qTNF6 and C1qTNF8 are homologous to and have similar domain structure as C1qTNF1. However, the intron-exon pattern of C1qTNF1 is somewhat different from those of C1qTNF6 and 8 (Figure 2). C1qTNF2 and 7 clearly fall into the same subfamily based on sequence homology, domain structure, and intron-exon pattern.

Murine C1qTNF3 was identified by suppression subtractive hybridization between TGF-β1 treated and untreated cells, and was also named CORS26 for "collagenous repeat-containing sequence of 26 kDa" (Maeda *et al., J. Biol. Chem.* 276:3628-3634 (2001) herein incorporated by reference in its entirety). It is expressed mainly in rib growth plate cartilage and kidney, and therefore may play a role in skeletal development (Maeda *et al.*, 2001, *supra*). It is also expressed in differentiated adipocytes (Schaffler *et al., Biochim. Biophys. Acta* 1628:64-70 (2003) herein incorporated by reference in its entirety). C1qTNF3 is coded by 6 exons, by far the most in all C1q related proteins (Figure 2).

C1qTNF4 is the only C1q related protein that contains more than one C1q domain. In addition, it is the only protein coded by a single exon. C1qTNF5 was recently identified as a gene associated with late-onset retinal degeneration (Hayward *et al., Hum. Mol. Genet.* 12:2657-2667 (2003) herein incorporated by reference in its entirety). A mutation in the C1q domain causes high molecular weight aggregate formation which may be causative of the disease. C1qTNF5 is mainly expressed in retinal pigment epithelium, liver, lung, brain and placenta (Hayward *et al.*, 2003, *supra*).

The present invention relates to two C1qTNF-like polypeptides. The first C1qTNF-like CDCP polypeptide of the invention (SEQ ID NO: 59) is an approximately 289 amino acid protein with a predicted molecular mass of approximately 32 kD unglycosylated. The initial methionine starts at position 80 of SEQ ID NO: 58 and the putative stop codon begins at position

947 of SEQ ID NO: 58. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 59 shares 100% identity with human C1qTNF-7, gi:13994280 (SEQ ID NO: 75) over 289 amino acids of SEQ ID NO: 75 (see Figure 14).

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 59 revealed its structural homology to C1q and collagen domains (see Table 31). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 31

e-value	Score	Model	Description	Amino acid position
1.3e-05	25.8	Collagen	Collagen triple helix repeat (20 copies)	37-73
2e-11	47.0	Collagen	Collagen triple helix repeat (20 copies)	77-136
1.3e-40	148.4	C1q	C1q domain	149-273

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 59 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 32). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 32

Accession number	Name	Amino acid position
IPB000885B	Fibrillar collagen C-terminal domain	3-161
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	10-164
IPB001073B	Complement C1q protein	31-275
PR00007A	Complement C1q domain signature I	158-184
PR00007B	Complement C1q domain signature II	185-204
PR00007C	Complement C1q domain signature III	229-250
PR00007D	Complement C1q domain signature IV	264-274

A predicted approximately 16 residue signal peptide is encoded from approximately residue 1 to residue 16 of SEQ ID NO: 59. The extracellular portion is useful on its own. The signal peptide region was predicted using the Neural Network SignalP V1.1 program (Nielsen *et al.*, 1997, *supra*). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program. SEQ ID NO: 60 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 59.

The second C1qTNF-like CDCP polypeptide of the invention (SEQ ID NO: 63) is an approximately 259 amino acid protein with a predicted molecular mass of approximately 28 kD unglycosylated. The initial methionine starts at position 138 of SEQ ID NO: 62 and the putative stop codon begins at position 915 of SEQ ID NO: 62. Protein database searches with the

BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 63 shares 100% identity with human C1qTNF-6, gi 32967294 (SEQ ID NO: 76) over 259 amino acids of SEQ ID NO: 76 (see Figure 15).

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 63 revealed its structural homology to C1q and collagen domains (see Table 33). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 33

e-value	Score	Model	Description	Amino acid position
7.2e-07	28.1	Collagen	Collagen triple helix repeat (20 copies)	78-119
2.1e-11	44.4	C1q	C1q domain	126-254

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 63 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 34). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 34

Accession number	Name	Amino acid position
IPB000885B	Fibrillar collagen C-terminal domain	46-144
IPB001442A	C-terminal tandem repeated domain in type 4 procollagen	53-144
IPB001073B	Complement C1q protein	71-228
PR00007A	Complement C1q domain signature I	137-163

A predicted approximately 27 residue signal peptide is encoded from approximately amino acid 1 to 27 of SEQ ID NO: 63. The extracellular portion is useful on its own. The signal peptide region was predicted using the Neural Network SignalP V1.1 program (Nielsen *et al.*, 1997, *supra*). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program. SEQ ID NO: 64 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 63.

The present invention also provides two nucleotide variants, both of which, when transcribed and translated produce the polypeptide of SEQ ID NO: 63. The first variant is represented in the attached sequence listing as SEQ ID NO: 65 and encodes the polypeptide of SEQ ID NO: 63. The initial methionine starts at position 123 of SEQ ID NO: 65 and the putative stop codon begins at position 900 of SEQ ID NO: 65. The second nucleotide variant that encodes the polypeptide of SEQ ID NO: 63 is SEQ ID NO: 66. The initial methionine starts at position 123 of SEQ ID NO: 66 and the putative stop codon begins at position 900 of SEQ ID

position 123 of SEQ ID NO: 66 and the putative stop codon begins at position 900 of SEQ ID NO: 66. The three nucleotide sequences of SEQ ID NO: 62, 65, and 66 differ in the 5' and 3' untranslated regions (see Figures 16 and 17).

The C1qTNF-like CDCP polypeptides of the invention are expected to share the same properties and activities as other C1qTNF polypeptides. Therefore, the C1qTNF-like polypeptides of the invention are expected to be useful in treating, diagnosing, and/or ameliorating diseases and disorders involving cartilage and bone development, lipid metabolism, diabetes, glucose and blood sugar metabolism, retinal degeneration, and other ophthalmic diseases, cardiovascular diseases, and kidney diseases.

Hibernation Proteins

In mammals, only a limited number of species, especially certain small mammals (i.e. chipmunks and squirrels), express hibernation. Many non-hibernating mammals retain the genes; however the transcripts are not detected. Mammalian hibernation is considered to be a unique physiological adaptation that allows life to be sustained under extremely low body temperatures. During hibernation, the body temperature drops to below 10 or 5°C, the heart and breathing rates fall and the metabolic rate is reduced to only a few percent of the euthermic levels (Kojima *et al, Eur. J. Biochem.* 268:5997-6002 (2001) herein incorporated by reference in its entirety). The chipmunk hibernation-associated proteins, HP-20, 25, 27 and 55 form a 140 kD complex in plasma. The expression level of this complex is tightly associated with the hibernation status of the animal: it drops before hibernation starts and increases before hibernation ends. HP-20, 25 and 27 are homologous to each other and each contains a collagen-like region followed by a C-terminal C1q domain. These genes are present, but not expressed in a non-hibernating squirrel (Takamatsu *et al., Mol. Cell Biol.* 13:1516-1521 (1993) herein incorporated by reference in its entirety).

The invention also relates to a CDCP polypeptide (SEQ ID NO: 68) that is a human orthologs of the chipmunk hibernation proteins. SEQ ID NO: 68 is an approximately 191 amino acid protein with a predicted molecular mass of approximately 21 kD unglycosylated. The initial methionine starts at position 44 of SEQ ID NO: 67 and the putative stop codon begins at position 617 of SEQ ID NO: 67. Protein database searches with the BLASTP algorithm (Altschul *et al.*, 1993, *supra*; and Altschul *et al.*, 1990, *supra*) indicate that SEQ ID NO: 68 shares 50% identity and 66% similarity with chipmunk HP-20 precursor, gi:1170339 (SEQ ID NO: 79) over 153 amino acids of SEQ ID NO: 79 (see Figure 18).

Using the Pfam software program (Sonnhammer *et al.*, 1998, *supra*), the CDCP polypeptide of SEQ ID NO: 68 revealed its structural homology to C1q domains (see Table 35). The results describe e-value, score, model, description, and amino acid position of the domain in the full-length protein.

Table 35

e-value	Score	Model	Description	Amino acid position
1.0e-42	152.1	C1q	C1q domain	47-173

Using the eMATRIX software package (Stanford University, Stanford, CA) (Wu *et al.*, 1999, *supra*), the CDCP polypeptide of SEQ ID NO: 68 was determined to have the following eMATRIX domain hits with e-values less than 1e-07 (see Table 36). The results describe: Accession number, name, and the position of the domain in the full-length protein.

Table 36

Accession number	Name	Amino acid position
PR00007A	Complement C1q domain signature I	56-82
IPB001073B	Complement C1q protein	63-151
PR00007B	Complement C1q domain signature II	83-102
PR00007C	Complement C1q domain signature III	132-153

A predicted approximately 24 residue signal peptide is encoded from approximately residue 1 to residue 24 of SEQ ID NO: 68. The extracellular portion is useful on its own. The signal peptide region was predicted using the Neural Network SignalP V1.1 program (Nielsen *et al.*, 1997, *supra*). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program. SEQ ID NO: 69 is the resulting peptide when the signal peptide is removed from SEQ ID NO: 68.

The hibernation protein-like CDCP polypeptide of the invention is expected to have similar properties and activities as the hibernation proteins. It is expected that SEQ ID NO: 68 will be useful in modulating body temperature, heart and breathing rates and metabolic rates. SEQ ID NO: 68 may be useful in treating hypothermia, frost bite, fat metabolism, and the like.

Expression of human CDCP proteins

About half of the 31 human CDCP proteins have reported spatial and/or temporal expression patterns. Several were reviewed previously (Kishore and Reid, *Immunopharmacology* 49:159-170 (2000) herein incorporated by reference in its entirety.) Most of them are expressed highly specifically correlating very well with their specific functions.

The following proteins show very strict tissue-specific expressions. Adiponectin is expressed exclusively in adipose tissue. COL10A1 is expressed specifically in hypertrophic

chondrocytes during endochondral ossification (Thomas *et al., Biochem. Soc. Trans.* 19:804-808 (1991) herein incorporated by reference in its entirety). CBLN1 and 3 are expressed in adult cerebellum (Urade *et al.*, 1991, *supra*; Pang *et al.*, 2000, *supra*). CBLN2 is expressed in extracellular brain areas and in fetal brain (Wada and Ohtani, 1991, *supra*). CRF1 is expressed mainly in areas of the nervous system involved in motor function (Berube *et al.*, 1999, *supra*). EMILIN2 is mainly expressed in the cochlear basilar membrane (Amma *et al.*, 2003, *supra*). EMILIN3 is expressed exclusively on blood vessel endothelium (Christian *et al.*, 2001, *supra*). Multimerin is expressed in platelet α-granules and in vascular endothelium (Hayward *et al.*, 1995a, *supra*; Hayward *et al.*, 1995b, *supra*).

The rest of the characterized CDCP proteins show different tissue specificity: instead of being expressed in only one tissue or cell type, they are expressed in a few tissues or several distinct cell types. COL8A1 and 2 are expressed in corneal endothelium and are also present in vascular subendothelial matrices, heart liver, kidney, lung, and in some tumors (reviewed in Shuttleworth, *Int. J. Biochem. Cell Biol.* 29:1145-1148 (1997) herein incorporated by reference in its entirety). C1qTNF1 is predominantly expressed in heart, but also is expressed in endothelial and vascular smooth muscle cells (Innamorati *et al.*, 2002, *supra*). C1qTNF3 is expressed mainly in rib growth plate cartilage and kidney (Maeda *et al.*, 2001, *supra*) and also in differentiated adipocytes (Schaffler *et al.*, 2003, *supra*). C1qTNF5 is expressed in retinal pigment epithelium, liver, lung, brain, and placenta (Hayward *et al.*, 2003, *supra*). EMILIN1 is highly expressed in blood vessels, skin, heart, and lung (Doliana *et al.*, 1999, *supra*).

CDCP proteins in other species

To study the evolutionary development of the CDCP protein family, the C1q domainsfrom all known human C1q genes were used to BLAST against the genpept and genomic databases of various species. Sequences with significant hits ($S \geq 100$, $p \leq 10e\text{-}6$) were collected and then a similar search was performed recursively. Thus, the discovery of one CDCP protein in a species may eventually bring several distinct members that belong to the same subfamily; the newer members may have low homology to the original CDCP genes. A Hidden Markhov Model of the C1q domains from Pfam was also applied to those same databases or 6-frame translated databases of genomic sequences, with p=0.001 as a cut off. Finally, multiple HMM models trained on different sets of confirmed C1q domains were developed and applied to the same databases in a recursive fashion.

Of the 31 human CDCP genes reported herein, 29 orthologous genes in *Mus musculus* were identified. Mouse orthologs to AQL2 and C1QTNF8 were not found. Since the DNA

sequence for human AQL2 is nearly identical to AQL1, it appears likely to have arisen from a very recent gene duplication event.

CDCP proteins were identified in species ranging from *Macaca mulatta* (monkey) to *Strongylocentrotus purpuratus* (purple sea urchin). Five CDCP family members were identified in the sea urchin with BLASTp S-score against the human CDCP domains in the range of 75-80 and p-value of 9.0×10^{-5} to 1.0×10^{-5}. A comparison of these sea urchin proteins to human adiponectin reveals conservation of 5 to 7 of the 8 residues found invariant in the human CDCP family (see Figure 19A). A comparison of a ribbon model for one of the CDCP proteins in the sea urchin, Sp_C1qDC4, to the crystal structure of human adiponectin suggests that these conserved residues have side chains in the area of the core of the globular structure of the C1q domain (Figure 4). For Sp_C1qDC4, the only substitution seen in the residues corresponding to the 8 invariant residues seen in human substitutes a tyrosine for phenylalanine and is consistent with the proposed hydrophobic packing core.

In addition, three very distant CDCP proteins were detected in the bacterium *Bacillius cereus*. For example, GenBank Accession AAP09378 has a C1q.hmm hmmsearch score of 3.0 and a p-value of 10e-3. In addition to encoding a weak, but apparent, C1q domain, this CDCP protein, named BC_C1qDC3, has the associated GenBank annotation: "collagen triple helix repeat protein." It contains a Pfam-detectable collagen domain of 21 amino acids before the C1q domain (score 3.8, p-value 1.3). This pattern of C1q domain preceeded by a collagen domain is seen in many human CDCP proteins and serves well to support the suggestion that Bc_C1qDC3 is a CDCP protein. Bc_C1qDC1 (SEQ ID NO: 85) and Bc_C1qDC2 (SEQ ID NO: 86) are both annotated as hypothetical proteins in GenBank and are much more closely related to each other than to Bc_C1qDC3 (SEQ ID NO: 87). The alignment of these three *B. cereus* CDCPs with human adiponectin is shown in Figure 19B. Five of the 8 conserved residues among all human CDCPs are also conserved in the *B. cereus* proteins.

No CDCP proteins were detected in other sequenced bacterial species. Neither were they found in the sequenced genomes of *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* or *Drosophila melanogaster*.

Other features of C1q related proteins

Crystal structures of adiponectin, COL8A1, and COL10A1 clearly indicate that the C1q domain is a trimerization structural element. Most C1q-related proteins also consist of a collagen-like region, which also has a tendency to trimerize. The trimerization of C1q domains

is suggested to nucleate the triple helix formation of the collagen-like regions. Conversely, the triple helical collagen stalk may stabilize the C1q trimer. For example, the recombinant C1q domain of adiponectin exists as both monomer and trimer, whereas full-length recombinant adiponectin forms trimers and hexamers (Yamauchi *et al.*, 2002, *supra*). It is expected that most, if not all, of the C1q-related proteins exist as homo- or hetero-trimers or higher order oligomers.

The intron-exon patterns of C1q related proteins are also diverse, although most patterns are conserved within subfamilies. Whereas most of the C1q domains are coded by one exon, 11 of them (those of CBLN1 to 4, gliacolin1, gliacolin2, CRF1, CRF2, C1QTNF3, EMILIN1, and EMILIN2) are coded by more than one exon (Figure 2). Among those whose C1q domains are coded by more than one exon, no clear evolutionary relationship between subfamilies can be drawn from these 4 intron-exon patterns. For those whose C1q domains are coded by one exon, one particular pattern with 2 exons is common in proteins from different subfamilies, including C1QA to C, adiponectin, C1QTNF2, 5, and 7. A slightly different pattern is shared by short chain collagens and C1QTNF6 and C1QTNF8. Exon patterns of AQL1 and AQL2, otolin, and C1QTNF1 could be derived from the above 2 patterns respectively (Figure 2). Therefore these genes are likely more related in evolution history.

Among all 31 C1q related proteins, adiponectin is the only protein studied so far to clearly demonstrate the ability of triggering signal transduction events in the cell. Recently, two cell surface receptors (adipoR1 and adipoR2) of adiponectin were identified (Yamauchi *et al.*, *Nature* 423:762-769 (2003) herein incorporated by reference in its entirety). These proteins are highly related and belong to a newly identified 7-transmembrane receptor family named PAQR (Tang *et al.*, "PAQR Proteins: A Novel Membrane Receptor Family Defined by an Ancient 7-transmembrane Pass Motif," (2004) submitted; and co-owned U.S. Provisional Application 60/498,969). A total of 11 PAQR members are found in human and mouse

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide that retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "biologically active" or "biological activity" refers to the capability of the natural, recombinant or synthetic C1q domain-containing peptide, or any peptide thereof, to induce a specific biological response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms “complementary” or “complementarity” refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5’-AGT-3’ binds to the complementary sequence 3’-TCA-5’. Complementarity between two single-stranded molecules may be “partial” such that only some of the nucleic acids bind or it may be “complete” such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term “embryonic stem cells (ES)” refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term “germ line stem cells (GSCs)” refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term “primordial germ cells (PGCs)” refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves. The term “totipotent” refers to the capability of a cell to differentiate into all of the cell types of an adult organism. The term “pluripotent” refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The term "expression modulating fragment," EMF, means a series of nucleotides that modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs comprises nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or “nucleic acid” or “polynucleotide” or “oligonculeotide” are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic

origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences, A is adenine, C is cytosine, G is guanine, and T is thymine, while N is A, T, G, or C. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequence herein may be replaced with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or “probe” or “primer” are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to a portion of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from any of the nucleic acid sequences of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58. The sequence information can be a segment of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms “operably linked” or “operably associated” refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term “pluripotent” refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms “polypeptide” or “peptide” or “amino acid sequence” refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term “translated protein coding portion” means a sequence which encodes for the full length protein which may include any leader sequence or a processing sequence.

The term “mature protein coding sequence” refers to a sequence which encodes a peptide or protein without any leader/signal sequence. The “mature protein portion” refers to that portion of the protein without the leader/signal sequence. The peptide may have the leader sequences removed during processing in the cell or the protein may have been produced synthetically or using a polynucleotide only encoding for the mature protein coding sequence. It is contemplated that the mature protein portion may or may not include an initial methionine residue. The initial methionine is often removed during processing of the peptide.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or “analog”) refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. “Conservative” amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or “substantially purified” as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more

preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other components normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g., E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which

have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligonucleotides), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about

35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 90% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, and most preferably at least about 95% identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term “totipotent” refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated

with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

4.2 NUCLEIC ACIDS OF THE INVENTION

The invention is based on the discovery of novel C1q domain-containing polypeptides, the polynucleotides encoding the C1q domain-containing polypeptides and the use of these compositions for the diagnosis, treatment or prevention of cardiovascular diseases, diseases/disorders related to lipid metabolism, glucose or blood sugar metabolism, obesity, diabetes, stroke, kidney diseases/disorders, extracellular matrix-associated diseases/disorders, chondrodysplasia, cellular senescence, neurological diseases, cartilage and/or bone development, retinal degeneration, ophthalmic diseases, auditory disorders, balance, hypothermia, and body temperature regulation.

The isolated polynucleotides of the invention include, but are not limited to a polynucleotide comprising any of the nucleotide sequences of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58; a polynucleotide comprising the full length protein coding sequence of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 (for example coding for SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69); and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polypeptides of any one of SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58; (b) a polynucleotide encoding any one of the polypeptides of SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding,

extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include the entire coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more typically at least about 90%, 91%, 92%, 93%, or 94% and even more typically at least about 95%, 96%, 97%, 98% or 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to any one of the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor result for the nucleic acids of the present invention, including SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990))

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also

encodes proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that

differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences, coding for any one of SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression

vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable

markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intramuscular injection of the DNA. The nucleic acid

sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.2.1 ANTISENSE NUCLEIC ACIDS

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that can hybridize to, or are complementary to, the nucleic acid molecule comprising a CDCP nucleotide sequence, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire CDCP coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of CDCP or antisense nucleic acids complementary to a CDCP nucleic acid sequence of are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a CDCP protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "conceding region" of the coding strand of a nucleotide sequence encoding CDCP protein. The term "conceding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a CDCP protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of CDCP mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of CDCP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of CDCP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the

duplex formed between the antisense and sense nucleic acids (e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following section).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a CDCP polypeptide to thereby inhibit expression of the protein (e.g., by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an alpha-anomeric nucleic acid molecule. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual alpha-units, the strands run parallel to each other. See, e.g., Gaultier, et al., 1987. Nucl. Acids Res. 15: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (see, e.g., Inoue, et al. 1987. Nucl. Acids Res. 15: 6131-6148) or a chimeric RNA-DNA analogue (see, e.g., Inoue, et al., 1987. FEBS Lett. 215: 327-330.

4.2.2 RIBOZYMES AND PNA MOIETIES

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they can be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach 1988. Nature 334: 585-591) can be used to catalytically cleave C1q domain-containing mRNA transcripts to thereby inhibit translation of C1q domain-containing mRNA. A ribozyme having specificity for a C1q domain-containing-encoding nucleic acid can be designed based upon the nucleotide sequence of a C1q domain-containing cDNA disclosed herein. For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a C1q domain-containing-encoding mRNA. See, e.g., U.S. Patent 4,987,071 to Cech, et al. and U.S. Patent 5,116,742 to Cech, et al. Stem cell growth factor-like mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel et al., (1993) Science 261:1411-1418.

Alternatively, C1q domain-containing gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the C1q domain-containing nucleic acid (e.g., the C1q domain-containing promoter and/or enhancers) to form triple helical structures that prevent transcription of the C1q domain-containing gene in target cells. See, e.g., Helene, 1991. Anticancer Drug Des. 6: 569-84; Helene, et al. 1992. Ann. N.Y. Acad. Sci. 660: 27-36; Maher, 1992. Bioassays 14: 807-15.

In various embodiments, the CDCP nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. See, e.g., Hyrup, et al., 1996. Bioorg Med Chem 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (e.g., DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, et al., 1996. supra; Perry-O'Keefe, et al., 1996. Proc. Natl. Acad. Sci. USA 93: 14670-14675.

CDCP PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. CDCP PNAs can also be used, for example, in the analysis of single base pair mutations in a gene (e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (see, Hyrup, et al., 1996.supra); or as probes or primers for DNA sequence and hybridization (see, Hyrup, et al., 1996, supra; Perry-O'Keefe, et al., 1996. supra).

In another embodiment, CDCP PNAs can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, CDCP PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (e.g., RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (see, Hyrup, et al., 1996. supra). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, et al., 1996. Supra, et al., 1996. Nucl Acids Res 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. See, e.g., Mag, et al., 1989. Nucl Acid Res 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. See, e.g., Finn, et al., 1996. supra. Alternatively, chimeric molecules

can be synthesized with a 5' DNA segment and a 3' PNA segment. See, e.g., Petersen, et al., 1975. Bioorg. Med. Chem. Lett. 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across the cell membrane (see, e.g., Letsinger, et al., 1989. Proc. Natl. Acad. Sci. U.S.A. 86: 6553-6556; Lemaitre, et al., 1987. Proc. Natl. Acad. Sci. 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol, et al., 1988. BioTechniques 6:958-976) or intercalating agents (see, e.g., Zon, 1988. Pharm. Res. 5: 539-549). To this end, the oligonucleotide can be conjugated to another molecule, e.g., a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

4.3 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., Basic Methods in Molecular Biology (1986)). The host cells containing one of polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed

in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell tines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

A number of types of cells may act as suitable host cells for expression of the protein. Mammalian host cells include, for example, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from in vitro culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces strains*, Candida, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary

to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting, including polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.3.1 CHIMERIC AND FUSION PROTEINS

The invention also provides CDCP chimeric or fusion proteins. As used herein, a CDCP g "chimeric protein" or "fusion protein" comprises a CDCP polypeptide operatively linked to a non-CDCP polypeptide. A " CDCP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a CDCP protein, whereas a "non-CDCP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the CDCP protein, e.g., a protein that is different from the CDCP protein and that is derived from the same or a different organism. Within a CDCP fusion protein the CDCP polypeptide can correspond to all or a portion of a CDCP protein. In one embodiment, a CDCP fusion protein comprises at least one biologically active portion of a CDCP protein. In another embodiment, a CDCP fusion protein comprises at least two biologically active portions of a CDCP protein. In yet another embodiment, a CDCP fusion protein comprises at least three biologically active portions of a CDCP protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the CDCP polypeptide and the non-CDCP polypeptide are fused in-frame with one another. The non-CDCP polypeptide can be fused to the N-terminus or C-terminus of the CDCP polypeptide.

In one embodiment, the fusion protein is a GST-C1q domain-containing fusion protein in which the CDCP sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant CDCP polypeptides. In another embodiment, the fusion protein is a CDCP protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of CDCP can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a CDCP-immunoglobulin fusion protein in which the CDCP sequences are fused to sequences derived from a member of the immunoglobulin protein family. The CDCP-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an

interaction between a CDCP ligand and a CDCP protein on the surface of a cell, to thereby suppress CDCP-mediated signal transduction *in vivo*. The CDCP-immunoglobulin fusion proteins can be used to affect the bioavailability of a CDCP cognate ligand. Inhibition of the CDCP ligand/CDCP interaction can be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (e.g. promoting or inhibiting) cell survival. Moreover, the CDCP-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-CDCP antibodies in a subject, to purify CDCP ligands, and in screening assays to identify molecules that inhibit the interaction of CDCP with a CDCP ligand.

A CDCP chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, e.g., by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, e.g., Ausubel, et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A CDCP nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the CDCP protein.

4.4 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequence set forth as any one of SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 or (b) polynucleotides encoding any one of the amino acid

sequences set forth as SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more typically at least about 90%, 91%, 92%, 93%, or 94% and even more typically at least about 95%, 96%, 97%, 98% or 99%, most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which it is expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a

nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to,

immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.,* as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.,* U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved

systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). The polypeptides of the invention include CDCP analogs. This embraces fragments of CDCP polypeptide of the invention, as well CDCP polypeptides which comprise one or more amino acids deleted, inserted, or substituted. Also, analogs of the CDCP polypeptide of the invention embrace fusions of the CDCP polypeptides or modifications of the CDCP polypeptides, wherein the CDCP polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the CDCP polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to specific cell types, such as neurons, e.g., antibodies to central nervous system, or antibodies to receptor and ligands expressed on neuronal cells. Other moieties which may be fused to CDCP polypeptides include therapeutic agents which are used for treatment, for example anti-depressant drugs or other medications for neurological disorders. Also, CDCP polypeptides may be fused to neuron growth modulators, and other chemokines for targeted delivery.

4.4.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), the eMatrix software (Wu et al., J. Comp. Biol., vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, vol 4, pp. 202-209, herein incorporated by reference), the

GeneAtlas software (Molecular Simulations Inc. (MSI), San Diego, CA) (Sanchez and Sali (1998) Proc. Natl. Acad. Sci., 95, 13597-13602; Kitson DH et al, (2000) "Remote homology detection using structural modeling – an evaluation" Submitted; Fischer and Eisenberg (1996) Protein Sci. 5, 947-955), and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

4.5 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.,* inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory

element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.6 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies; of animals that fail to express functional C1q domain-containing polypeptide or that express a variant of C1q domain-containing polypeptide. Such animals are useful as models for studying the *in vivo* activities of C1q domain-containing polypeptide as well as for studying modulators of the C1q domain-containing polypeptide.

4.7 USES AND BIOLOGICAL ACTIVITY OF HUMAN C1Q DOMAIN-CONTAINING POLYPEPTIDES

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple

helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.7.1 RESEARCH USES AND UTILITIES

In addition to the therapeutic and diagnostic uses of the polypeptides and polynucleotides of the invention stated herein, the polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

4.7.2 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present

invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9-Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober,

Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.7.3 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* may maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected

with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. Furthermore, these cells can be cultured *in vitro* to form other differentiated cells, such as skin tissue that can be used for transplantation. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as

retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.7.4 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional colony stimulating factor activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

4.7.5 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies

resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pp. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

4.7.6 IMMUNE FUNCTION STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells

and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and

persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating

autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce

tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.7.7 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.

M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

4.7.8 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

4.7.9 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid

phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl,

Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

4.7.10 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

4.7.11 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the

diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate “hit” (or “lead”) to optimize the capacity of the “hit” to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding

molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

4.7.12 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.,* increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with a polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

4.7.13 LEUKEMIA

Leukemia and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

4.7.14 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin,

trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

4.7.15 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.7.16 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed *Mycobacterium tuberculosis* in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

4.7.17 METABOLIC DISORDERS

A polynucleotide and polypeptide of the invention may also be involved in the prevention, diagnosis and management of metabolic disorders involving carbohydrates, lipids, amino acids, vitamins etc., including but not limited to diabetes mellitus, obesity, aspartylglusomarinuria, carbohydrate deficient glycoprotein syndrome (CDGS), cystinosis, diabetes insipidus, Fabry, fatty acid metabolism disorders, galactosemia, Gaucher, glucose-6-phosphate dehydrogenase (G6PD), glutaric aciduria, Hurler, Hurler-Scheie, Hunter, hypophosphatemia, I-cell, Krabbe, lactic acidosis, long chain 3 hydroxyacyl CoA dehydrogenase deficiency (LCHAD), lysosomal storage diseases, mannosidosis, maple syrup urine, , Maroteaux-Lamy, metachromatic leukodystrophy, mitochondrial Morquio, mucopolysaccharidosis, neuro-metabolic, Niemann-Pick, organic acidemias, purine, phenylketonuria (PKU), Pompe, porphyria, pseudo-Hurler, pyruvate dehydrogenase deficiency,

Sandhoff, Sanfilippo, Scheie, Sly, Tay-Sachs, trimethylaminuria (Fish-Malodor syndrome), urea cycle conditions, vitamin D deficiency rickets and related complications involving different organs including but not limited to liver, heart, kidney, eye, brain, muscle development etc. Hereditary and/or environmental factors known in the art can predispose an individual to developing metabolic disorders and conditions resulting therefrom. Under these circumstances, it maybe beneficial to treat these individual with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing the disorder. Examples of such disorders include diabetes mellitus, obesity and cardiovascular disease. Further, polynucleotide sequences encoding the invention may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered expression of the polynucleotides of the invention. Such qualitative or quantitative methods are well known in the art.

4.7.18 CARDIOVASCULAR DISEASE AND THERAPY

Polypeptides and polynucleotides of the invention may also be involved in the prevention, diagnosis and management of cardiovascular disorders such as coronary artery disease, atherosclerosis and hyper- and hypolipoproteinemia, hypertension, angina pectoris, myocardial infarction, congestive heart failure, cardiac arrythmias including paroxysmal arrythmias, restenosis after angioplasty, aortic aneurysm and related complications involving various organs including but not limited to kidney, eye, brain, heart etc. Polypeptides of the invention may also have direct and indirect effects on myocardial contractility, electrical activity of the heart, atrial fibrillation, atrial fluter, anomalous atrio-ventricular pathways, sino-atrial dysfunction, vascular insufficiency and arterial embolism. Hereditary and/or environmental factors known in the art can predispose an individual to developing metabolic disorders and conditions resulting therefrom. Under these circumstances, it may be beneficial to treat these individual with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing the disorder. Examples of such disorders include but are not limited to coronary artery disease, atherosclerosis, hyper- and hypolipoproteinemia, hypertension, angina pectoris, myocardial infarction, cardiac arrythmias including paroxysmal arrythmias, diabetes mellitus, inflammatory glomerulonephritis, ischemic renal failure, extracellular matrix accumulation, fibrosis, hypertension, coronary vasoconstriction, ischemic heart disease, and lesions occurring in brain disorders such as stroke, trauma, infarcts, aneurysms.

The polynucleotide sequences encoding the invention may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in

dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered expression of the polynucleotides of the invention. Such qualitative or quantitative methods are well known in the art.

4.8 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

4.8.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the CDCP polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of CDCP polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 mg/kg of patient body weight. For parenteral administration, C1q domain-containing polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

4.9 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered

to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest

edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co- administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

4.9.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically,

for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

4.9.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other

glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically

acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for

the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl

cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredient of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredient of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a

mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

4.9.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted

individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

4.9.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.10 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from

humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 4-5, 7-8, 19-20, 24-25, 27-28, 32, 34-35, 38-39, 41-42, 46, 48, 51, 55, 59-60, or 68-69, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828 (1981); Kyte and Doolittle, *J. Mol. Biol.* 157: 105-142 (1982), each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow *et al.* (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The

present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. *et al.,* Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

4.10.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface-active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, *etc.*), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques,

such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

4.10.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized *in vitro*.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur *et al.*, Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of

monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, *Nature* 368:812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

4.10.3 HUMANIZED ANTIBODIES

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones *et al.*, *Nature*, 321:522-525 (1986); Riechmann, *et al., Nature*, 332:323-327 (1988); Verhoeyen, *et al., Science*, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones *et al.*, 1986; Riechmann *et al.*, 1988; and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)).

4.10.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, *et al.*, *Immunol Today* 4: 72 (1983)) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, *et al.*, *Proc Natl Acad Sci USA* 80: 2026-2030 (1983)) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks *et al., J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks *et al.* (*Bio/Technology* 10,:779-783 (1992)); Lonberg *et al.* (*Nature* 368:856-859 (1994)); Morrison (*Nature* 368:812-13 (1994)); Fishwild *et al*,(*Nature Biotechnology*, 14:845-51 (1996)); Neuberger (*Nature Biotechnology*, 14:826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13:65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™

as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

4.10.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, *et al.*, *Science* 246:1275-1281 (1989)) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen

may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

4.10.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, *EMBO J.*, 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh *et al., Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side

chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.* alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full-length antibodies or antibody fragments (*e.g.* $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan *et al., Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby *et al., J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling *in vitro* to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny *et al., J. Immunol.* 148:1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger *et al., Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of

another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber *et al., J. Immunol*. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt *et al., J. Immunol*. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

4.10.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

4.10.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron *et al., J. Exp Med.*, 176: 1191-

1195 (1992) and Shopes, *J. Immunol.*, 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff *et al. Cancer Research*, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson *et al., Anti-Cancer Drug Design*, 3: 219-230 (1989).

4.10.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta *et al., Science*, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation

using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is in turn conjugated to a cytotoxic agent.

4.11 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

4.13 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period

sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test

sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.14 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.15 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs

of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

4.16 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from of any of the nucleotide sequences SEQ ID NO: 1-3, 6, 18, 21-23, 26, 29-31, 33, 36-37, 40, 43, 44-45, 47, 49-50, 52-54, or 56-58 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides

additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

4.17 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, 1990 J. Clin Microbiol 28(6) 1462-72); using UV light (Nagata *et al.,* 1985; Dahlen *et al.,* 1987; Morrissey & Collins, Mol. Cell Probes 1989 3(2)

189-207) or by covalent binding of base modified DNA (Keller *et al.,* 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci USA 91(8) 3072-6 describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.,* (1991) Anal Biochem 198(1) 138-42.

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., 1991). In this technology, a phosphoramidate bond is employed (Chu et al., 1983 Nucleic Acids 11(18) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm$_7$), is then added to a final concentration of 10 mM 1-MeIm$_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-MeIm$_7$, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal Biochem 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.,* (1994) Proc. Natl. Acad. Sci USA 91(11) 5022-6. These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.18 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6. In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*JI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*JI**), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*JI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*JI** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled

quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.19 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

5. EXAMPLES

5.1 EXAMPLE 1

Generation of the complete set of human C1q-related proteins

To obtain a complete set of human C1q domain-containing proteins, a two-step recursive search was performed using adiponectin as the initial query. First, all of the homologous proteins from both the public and the Nuvelo proprietary full-length protein databases were collected. Then these proteins were used to search for new genes from the public and Nuvelo proprietary EST sequences and human genomic sequences. All genes were then examined for editing quality and in three cases (C1qDC1, C1qTNF6, and otolin) revised to new versions based on EST and genomic sequence information from human, mouse and other species. They were also checked for the presence of the C1q domain. The final list contains 31 proteins (Table 1). The cloning of 19 of these proteins or their orthologs in other species has been described in the literature. Twenty-five (25) and 26 of the C1q-related proteins match to human and mouse UniGene clusters, respectively (Table 1). All but one (C1qTNF8) of the C1q-related proteins have at least partial EST support from either human or mouse. All but two of the proteins (AQL2 and C1qTNF8) have mouse orthologs.

Nomenclature

Twenty (20) proteins on this list have HUGO official names and symbols (Table 1). Most of the HUGO names/symbols are used herein. In two cases, a different name is used instead of the HUGO name/symbol: 1) the other name is more popular in the literature (*i.e.* C1qC); and 2) the other name is more appropriately named as a subfamily member (*i.e.* CBLN4). For the rest of the 11 proteins, most of the existing names that were published in the literature or in GenBank were maintained. A few were renamed to better represent their familial relationship with others. Specifically, adiponectin was chosen instead of ACRP30, adipoQ, APM1, or GBP28 since it is the most commonly used name: AQL1 and 2 (adipoQ-like 1 and 2) were so named because they are two closely related proteins with best homology to adiponectin; C1QTNF8, a predicted gene, was renamed from "similar to C1QTNF6" to keep with the naming convention of "C1q and TNF-related proteins"; CRF (C1q-related factor) and its closely related protein "similar to CRF" were renamed to CRF1 and CRF2, respectively, to reflect the relationship of these two proteins; similarly, gliacolin and "gliacolin-like protein" were renamed

to gliacolin1 and gliacolin2, respectively; one protein was named otolin because it was believed to be the orthologs of salmon otolin.

Therefore, the following is the complete set of human C1q domain-containing proteins: adiponectin, AQL1, AQL2, C1qA-C, C1qDC1, C1qTNF1-8, CBLN1-4, COL8A1, COL8A2, COL10A1, CRF1, CRF2, EMILIN1-3, gliacolin1-2, multimerin, and otolin. Most closely related proteins bear the same name with a different numeric suffix. The only exception is the C1QTNF proteins which do not belong to a distinct subfamily.

5.2 EXAMPLE 2

General Bioinformatics Tools

General bioinformatics tools used for sequence analysis, such as signal peptide prediction, Pfam domain searches, pair-wise and multiple sequence alignment, and phylogenetic tree generation, were the same as described in (Tang *et al.*, 2004, *supra*; Tang *et al.*, "TAFA: A Novel Secreted Family with Homology to CC-Chemokines," *Genomics,* In press (2004), herein incorporated by reference in their entirety). Chromosomal location and human-mouse synteny analysis were performed using the UCSC Genome Browser (University of California, Santa Cruz) with April 2003 release for human and February 2003 release for mouse. Fugu genomic sequence information was obtained from the JGI Fugo Genome Project v3.0 site (Joint Genome Institute; Aparicio *et al.*, *Science* 297:1301-1310 (2002) herein incorporated by reference in its entirety).

Search for the Complete Set of Human C1q Domain-containing Proteins

The search for human CDCP genes was begun by taking the C1q domains from then-known human CDCP proteins, including human adiponectin, C1qA-c, *etc.* and performing an initial BLASTP search for homologous sequences from the primate subsection of GenBank nr (gbpri). Human sequences that scored S≥100 were evaluated for the presence of a C1q domain and collected. This search was repeated recursively with newly identified homologous sequences until no additional paralogs were identified. This approach identified all known CDCP genes in the public databases at that time.

To discover novel CDCP proteins within the human genome, the C1q domains from all known CDCP proteins were used as query to search for tBLASTn hits in human EST (dbEST and private) and genomic sequences with a BLAST cutoff of 70. Subsequently, these new hits were attempted to assemble into new genes as described previously (Tang *et al.*, 2004a, *supra*; Tang *et al.*, 2004b, *supra*). All collected genes were examined for editing quality, and in several

cases they were revised to new versions based on EST and genomic sequence information from human, mouse and other species. These genes were also confirmed by the presence of the C1q domain.

In attempting to identify additional human CDCP genes, the Pfam model C1q domain was used to search against the 6-frame translated dbEST databases from public and an in-house human EST database, the Derwent Geneseq nucleotide database, and also the human genomic data from GenBank. Applicants assembled a Hidden Markov model for the C1q family using the HMMER tool *hmmbuild* (Durbin *et al.*, "Biological sequence analysis: probabilistic models of proteins and nucleic acids," Cambridge: Cambridge University Press (1998) herein incorporated by reference in its entirety) and various combinations of known C1q domains from multiple species, and then used this model to search 6-frame translated EST databases, cDNAs and the human genome.

Human-Mouse Orthology

Mouse orthologs of human C1q-related proteins were identified using BLASTp to search the GenBank genpept database (genbank release 135). Orthology was assigned initially if both genes scored as the top BLASTp hit in a crosswise comparison (human gene vs. mouse nr, mouse gene vs. human nr). For some C1q mRNA sequences, mouse orthologs were not present in GenBank and so the human sequences were used to search the mouse genome with the UCSC Genome Browser (University of California, Santa Cruz); corresponding mouse genes were thus predicted based on the human protein sequences, EST and genomic sequence information. Two human C1q proteins (AQL2 and C1qTNF8) do not have apparent mouse orthologs. Human and mouse orthologs were aligned pairwise and were then re-examined for editing quality in the revision of several of the mouse sequences.

Human-mouse synteny was determined by mapping each pair of orthologs to their corresponding genomes with the UCSC Genome Browser and comparing their flanking genes. It is considered syntenic if orthologous gene(s) is identified in neighboring genes at least on one side of the query gene, since these gene pairs are the best BLAST hits for each other in the two genomes.

In addition to orthologs of the human C1q-related proteins, other possible mouse C1q domain-containing proteins were searched by tBLASTn against mouse genomic sequences using C1q domains of mouse orthologs of human C1q-related proteins with a cutoff score of 100. No new C1q domain-containing proteins were found.

Structural Modeling

Three-dimensional structural models of the AQL1 and C1qTNF7 proteins were generated using the GeneAtlas™ software package (Accelrys, San Diego, CA 1999). These models were predicted based on a search of 4250 non-redundant Protein Data Bank structures using a PSI-BLAST multiple alignment sequence profile-based searching method (Meyers and Miller, *Comput. Appl. Biosci.* 4:11-17 (1988) herein incorporated by reference in its entirety) and high throughput homology modeling, an automated sequence and structure searching procedure (Sali and Overington, *Protein Sci.* 3:1582-1596 (1994) herein incorporated by reference in its entirety). The known crystal structure of adiponectin (Shapiro and Scherer, 1998, *supra*) was identified as the best fit structure and was used as a template for structural overlays using Profiles-3D, a threading program that measures the compatibility of the protein model with its sequence using a 3-D profile. Using defined parameters, Profiles-3D computes a score for the model normalized by the length of the amino acid sequence.

AQL1 and AQL2 Genes in Other Primates

To investigate the presence of AQL1 and AQL2 genes in other primates, tBLASTn searches using AQL1 and AQL2 against EST and genomic sequences in NCBI were performed. These initial efforts yielded no orthologous sequence from other primates. Therefore, sequencing traces of macaca (*Macaca mulatta*) and chimpanzee (*Pan troglodytes*) were downloaded from NCBI and tBLASTn searches were performed against them with the AQL1 and AQL2 sequences. No orthologs was found from macaca traces probably due to the small amount of sequences available. Several orthologous sequences from chimpanzee were identified, and each of them shares 95% or higher sequence identity with AQL proteins. It appears that there are two slightly different versions of AQL orthologs in chimpanzee, one (represented by the trace name G591P68203FC1.T0) is closer to AQL1 than AQL2, the other (represented by the trace name G591P56972RE2.T0) is different from the first one with 5 different residues in a ~170 amino acid region. However, no sufficient data is available to determine whether or not this sequence is the orthologs of AQL2 since the trace sequence only covers part of the gene.

Pseudogenes

Several pseudogenes and partial pseudogenes are found in the human genome. One processed pseudogene, located at chromosome 6q25.1, shares good homology with EMILIN3, but lacks the N-terminal 157 amino acids (out of 946 amino acids) and contains several stop

codons and frameshifts. Another partial pseudogene is located at chromosome 19q13.32, and is homologous to the C1q domain region of EMILIN1 with a frameshift. Interestingly, at least three processed pseudogenes and many fragments homologous to C1q-related proteins are clustered in a ~250 kb region at chromosome 22q12.3, and no other genes are found in this region. These 3 pseudogenes, like the one homologous to EMILIN3, also lack the N-terminal region (~45 amino acids). Remarkably, these pseudogenes and fragments are most homologous to chipmunk hibernation proteins HP-20, 25, and 27. Therefore, this chromosomal region appears to be evolved from the same ancestor genes as those hibernation genes. However, this region is not found in mouse, probably due to the loss of this region in mouse during evolution, or this region of the mouse genome has not been sequenced.

5.3 EXAMPLE 3

Isolation of SEQ ID NO: 1, 21, 29, 36, 43, 52, and 56 from a cDNA Libraries of Human Cells

The novel nucleic acids of SEQ ID NO: 1, 21, 29, 36, 43, 52, and 56 were obtained from various human cDNA libraries using standard PCR, sequencing by hybridization sequence signature analysis, and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for vector sequences flanking the inserts. These samples were spotted onto nylon membranes and interrogated with oligonucleotide probes to give sequence signatures. The clones were clustered into groups of similar or identical sequences, and single representative clones were selected from each group for gel sequencing. The 5' sequence of the amplified inserts were then deduced using the reverse M13 sequencing primer in a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single-pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer. These inserts was identified as a novel sequence not previously obtained from this library and not previously reported in public databases. These sequences are designated as SEQ ID NO: 1, 21, 29, 36, 43, 52, and 56 in the attached sequence listing.

5.4 EXAMPLE 4

Assemblage of SEQ ID NO: 2, 22, 44, 53, or 57

The novel nucleic acids (SEQ ID NO: 2, 22, 44, 53, or 57) of the invention were assembled from sequences that were obtained from various cDNA libraries by methods described in Example 1 above, and in some cases obtained from one or more public databases. The final sequence was assembled using the EST sequence as seed. Then a recursive algorithm was used to extend the seed

into an extended assemblage, by pulling additional sequences from different databases (i.e. Nuvelo's database containing EST sequences, dbEST, gb pri, and UniGene) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

The nearest neighbor results for the assembled contigs were obtained by a FASTA search against Genpept, using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor results showed the closest homologue for each assemblage from Genpept (and contain the translated amino acid sequences for which the assemblages encodes). The nearest neighbor results are set forth in Table 37 below:

Table 37

SEQ ID NO:	Accession No.	Description	Smith-Waterman Score	% Identity
2	L23982	Homo sapiens collagen type VII	521	46.226
44	U27838	Mus musculus glycosyl-phosphatidyl-inositol-anchored protein homolog	418	29.216
53	AF095737	Homo sapiens unknown	366	68.085
57	X53556	Bos taurus type X collagen	657	42.963

The predicted amino acid sequences for SEQ ID NO: 2, 22, 44, 53, or 57were obtained by using a software program called FASTY (University of Virginia) which selects a polypeptide based on a comparison of translated novel polynucleotide to known polynucleotides (W.R. Pearson, Methods in Enzymology, 183:63-98 (1990), incorporated herein by reference). For SEQ ID NO: 2, 22, 44, 53, or 57, the predicted start and stop nucleotide locations are listed in Table 38:

Table 38

SEQ ID NO:	Predicted beginning nucleotide location corresponding to first amino acid residue of amino acid sequence	Predicted end nucleotide location corresponding to first amino acid residue of amino acid sequence
2	739	794
22	202	2471
44	3	2456
53	2471	2985
57	142	1058

5.5 EXAMPLE 5

Assemblage of SEQ ID NO: 3, 6, 9, 18, 23, 45, 48, or 58

The novel nucleic acids (SEQ ID NO: 3, 6, 9, 18, 23, 45, 48, or 58) of the invention were assembled from sequences that were obtained from cDNA libraries by methods described in Example 1 above, and in some cases obtained from one or more public databases. The final sequences were assembled using the EST sequences as seed. Then a recursive algorithm was used to extend the seed into an extended assemblage, by pulling additional sequences from different databases (i.e. Nuvelo's database containing EST sequences, dbEST, gb pri, and UniGene) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full-length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect sop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and BLAST against Genbank (i.e. dbEST, gb pri, UniGene, Genpept). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Nuvelo, Inc.). The full-length nucleotide sequences are shown in the Sequence Listing as SEQ ID NO: 3, 6, 9, 18, 23, 45, 47, or 58; and the full-length amino acid sequences are shown in the sequence listing as SEQ ID NO: 4, 7, 9, 19, 24, 46, 48, or 59.

Further annotation of SEQ ID NO: 45 or 47 can be found in U.S. patent application Serial No. 09/598,075 filed June 20, 2000 (attorney docket no. 787); herein incorporated by reference in its entirety.

Further annotation of SEQ ID NO: 23 can be found in U.S. patent application Serial No. 09/620,312 filed July 19, 2000 (attorney docket no. 784); herein incorporated by reference in its entirety.

Further annotation of SEQ ID NO: 58 can be found in U.S. patent application Serial No. 09/728,952 filed November 30, 2000 (attorney docket no. 799); herein incorporated by reference in its entirety.

Further annotation of SEQ ID NO: SEQ ID NO: 3, 6, 9 or 18 can be found in U.S. Provisional patent application Serial No. 60/306971 filed July 21, 2001 (attorney docket no. 805); herein incorporated by reference in its entirety.

5.6 EXAMPLE 6

Assemblage of SEQ ID NO: 31, 33, 37, 40 or 54

The novel nucleic acids (SEQ ID NO: 31, 33, 37, 40 or 54) of the invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases obtained from one or more public databases. The final sequence was assembled using the EST sequences as seed. Then a recursive algorithm was used to extend the seed into an extended assemblage, by pulling additional sequences from different databases (i.e. Nuvelo's database containing EST sequences, dbEST, gb pri, and UniGene) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full-length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect sop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and BLAST against Genbank (i.e. dbEST, gb pri, UniGene, Genpept). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Nuvelo, Inc.). The full-length nucleotide sequences are shown in the Sequence Listing as SEQ ID NO: 31, 33, 37, 40 or 54; and the full-length amino acid sequences are shown in the sequence listing as SEQ ID NO: 32, 34, 38, 41, or 55.

5.7 EXAMPLE 7

Tissue Expression Analysis and Chromosomal Localization of Full-length Polynucleotides of the Invention

By checking the Nuvelo proprietary database established from screening by hybridization, SEQ ID NO: 45 or 47 was found to be expressed in following human tissue/cell cDNA (see Table 39):

Table 39

Library Name	No. of Positive Clones	Total No. of Clones in the Library	Tissue Origin
BMD001	13	342599	bone marrow
ABD003	3	83268	adult brain
FLS001	30	555770	fetal liver-spleen

Library Name	No. of Positive Clones	Total No. of Clones in the Library	Tissue Origin
AKD001	5	176438	adult kidney
LUC001	5	210372	leukocytes
ATS001	2	26744	testis
AKT002	7	149669	adult kidney
AOV001	22	259409	adult ovary
IB2002	21	265743	infant brain
LGT002	7	158948	lung tumor
HFB001	5	74494	fetal brain
IBS001	3	33191	infant brain
LPC001	8	97546	lymphocyte
PIT004	5	120274	pituitary gland
SPC001	2	61905	whole organ
THM001	4	113947	thymus
THR001	2	124110	thyroid gland
ADR002	5	90185	adrenal gland
CVX001	7	125473	cervix
THA002	1	32817	thalamus
FUC001	1	125570	umbilical cord
SIN001	2	142562	whole organ
ABR001	3	30163	adult brain
FLG001	2	28154	whole organ
BLD001	3	29386	bladder
FSK001	5	127263	fetal skin
CLN001	3	28708	colon
REC001	1	28337	rectum
SPLc01	2	110573	spleen
FLG003	1	27360	fetal lung
NTU001	4	37055	neuronal cells
NTD001	5	35080	induced neuronal cells
NTR001	3	34629	retinoic acid-induced neuronal cells
ABR006	1	108204	adult brain
FBR004	1	27560	fetal brain
FBR006	8	151893	fetal brain
ABR008	14	145661	adult brain
FLS002	58	709733	fetal liver-spleen
IB2003	14	201294	infant brain
ADP001	2	37287	cultured preadipocytes
ADP002	1	32855	cultured preadipocytes
FLV002	2	32865	fetal liver
BMD002	1	75816	bone marrow
DIA002	1	40119	diaphragm
FLV004	3	74491	fetal liver
FKD002	1	33111	fetal kidney
FSK002	1	72628	fetal skin
FLS003	9	187791	fetal liver-spleen
HMP001	3	71425	macrophage
FLG004	1	41090	fetal lung
BMD008	1	44770	bone marrow
DGD001	1	91971	lymphocyte
DGD004	1	91423	lymphocytes
STM001	2	181899	bone marrow
OBE01	3	132217	adipocytes

SEQ ID NO: 45 or 47 were further analyzed for their presence in the public dbEST database and their tissue source. SEQ ID NO: 45 or 47 were found to be expressed in following tissues: Gessler Wilms tumor, colon, Stratagene hNT neuron, Fibroblasts, senescent, Stratagene endothelial cell 937223, Soares breast 2NbHBst, Stratagene lung carcinoma 937218, Soares fetal liver spleen 1NFLS, Soares_parathyroid_tumor_NbHPA, total brain, Soares_NhHMPu_S1, Soares_fetal_heart_NbHH19W, liver, Soares infant brain 1NIB, Jurkat T-cells, cochlea, Ovary, and Testis tumor.

The gene corresponding to SEQ ID NO: 45 or 47 was mapped to human chromosome 12p11-37.2 by BLAST analysis with human genome sequences.

By checking the Nuvelo proprietary database established from screening by hybridization, SEQ ID NO: 23 was found to be expressed in following human tissue/cell cDNA (see Table 40):

Table 40

Library Name	No. of Positive Clones	Total No. of Clones in the Library	Tissue Origin
LGT002	5	158948	lung tumor
MMG001	1	131991	mammary gland
PIT004	1	120274	pituitary gland
THR001	5	124110	thyroid gland
ADR002	2	90185	adrenal gland
TRC001	1	23820	trachea
FUC001	17	125570	umbilical cord
FLG001	1	28154	whole organ
FSK001	1	127263	fetal skin
ADP001	1	37287	adipocytes
ADP002	7	32855	adipocytes
PLA003	1	80877	placenta
FKD002	1	33111	fetal kidney
FSK002	1	72628	fetal skin
FHR001	2	108446	fetal heart
FLG004	1	41090	fetal lung
OBE01	5	132217	adipocytes

SEQ ID NO: 23 was further analyzed for their presence in the public dbEST database and their tissue source. SEQ ID NO: 23 was found to be expressed in following tissues: Bone, poorly differentiated adeno, Fibroblasts, senescent, melanocyte, colon tumor RER+, Soares_NhHMPu_S1, bone marrow stroma, 2 pooled tumors (clear cell, Soares ovary tumor NbHOT, cochlea.

The gene corresponding to SEQ ID NO: 23 was mapped to chromosome 3 by BLAST analysis with human genome sequences.

By checking the Nuvelo proprietary database established from screening by hybridization, SEQ ID NO: 58 was found to be expressed in following human tissue/cell cDNA (see Table 41):

Table 41

Library Name	No. f Positive Clones	Total No. of Clones in the Library	Tissue Origin
FLS001	1	555770	fetal liver-spleen
AKD001	3	176438	adult kidney
AOV001	9	259409	adult ovary
CVX001	2	125473	adult cervix
FLG001	1	28154	fetal lung
SPLc01	1	110573	spleen
FKD002	2	33111	fetal kidney

SEQ ID NO: 58 was further analyzed for their presence in the public dbEST database and their tissue source. SEQ ID NO: 58 was found to be expressed in following tissues: Soares_NhHMPu_S1, NCI_CGAP_Sub6.

The gene corresponding to SEQ ID NO: 58 was mapped to human chromosome 4 by BLAST analysis with human genome sequences.

By checking the Nuvelo proprietary database established from screening by hybridization, SEQ ID NO: 3, 6, 9, or 18 was found to be expressed in following human tissue/cell cDNA (see Table 42):

Table 42

Library Name	No. of Positive Clones	Total No. of Clones in the Library	Tissue Origin
FLS001	1	555770	fetal liver-spleen
FMS001	1	32743	Fetal muscle
FSK001	1	127263	Fetal skin
FMS002	6	40223	Fetal muscle
FHR001	4	108446	Fetal heart

SEQ ID NO: 3, 6, 9, or 18 was further analyzed for their presence in the public dbEST database and their tissue source. SEQ ID NO: 3, 6, 9, or 18 was found to be expressed in following tissues:

HEMBB1, head_normal, MAGE resequences, MAGM, bone marrow, larynx tumor, high grade preneoplastic lesion, NCI_CGAP_Sub7, NIH_MGC_87, NIH_MGC_91, Soares_NFL_T_GBC_S1, Soares_testis_NHT.

The gene corresponding to SEQ ID NO: 3, 6, 9, or 18 was mapped to human chromosome 13 by BLAST analysis with human genome sequences.

By checking the Nuvelo proprietary database established from screening by hybridization, SEQ ID NO: 31 or 33 was found to be expressed in following human tissue/cell cDNA (see Table 43):

Table 43

Library Name	No. of Positive Clones	Total No. of Clones in the Library	Tissue Origin
FLS001	1	555770	fetal liver-spleen
LUC001	1	210372	leukocytes
AKT002	1	149669	adult kidney
IB2002	2	265743	infant brain
HFB001	3	74494	fetal brain
SPC001	1	61905	whole organ
NTR001	1	34629	retinoic acid-induced neuronal cells
STM001	1	181899	bone marrow

SEQ ID NO: 31 or 33 was further analyzed for their presence in the public dbEST database and their tissue source. SEQ ID NO: 31 or 33 was found to be expressed in following tissues: 2 pooled tumors, HTC, and Soares fetal liver spleen 1NFLS S1.

The gene corresponding to SEQ ID NO: 31 or 33 was mapped to human chromosome 18 by BLAST analysis with human genome sequences.

By checking the Nuvelo proprietary database established from screening by hybridization, SEQ ID NO: 54 is found to be expressed in following human tissue/cell cDNA (see Table 44):

Table 44

Library Name	No. of Positive Clones	Total No. of Clones in the Library	Tissue Origin
BMD001	2	342599	bone marrow
ABD003	16	83268	adult brain
FLS001	2	555770	fetal liver-spleen
AKD001	2	176438	adult kidney
LUC001	3	210372	leukocytes
LUC003	3	30296	leukocytes
ALV001	1	30866	young liver
ATS001	1	26744	testis
ASP001	1	32114	adult spleen
APL001	1	31936	placenta
ABT004	732	31910	adult brain
AKT002	2	149669	adult kidney
ALV002	10	144402	adult liver
AOV001	5	259409	ovary
IB2002	1276	265743	infant brain
LGT002	16	158948	adult lung
MMG001	8	131991	mammary gland
HFB001	38	74494	fetal brain
FBT002	1	35745	fetal brain
IBM002	99	13952	infant brain
IBS001	182	33191	infant brain

Library Name	N . of Positive Clones	Total No. of Clones in the Library	Tissue Origin
LPC001	3	97546	lymphocyte
PIT004	3	120274	pituitary gland
SPC001	1705	61905	whole organ
THR001	1	124110	thyroid gland
MEL004	17	30503	melanoma
ADR002	3	90185	adrenal gland
CVX001	4	125473	cervix
PRT001	2	28649	whole organ
THA002	591	32817	thalamus
TRC001	1	23820	trachea
FBR001	1	28664	fetal brain
FUC001	8	125570	umbilical cord
SKM001	1	28327	whole organ
SIN001	6	142562	whole organ
ABR001	241	30163	adult brain
FLG001	2	28154	whole organ
BLD001	43	29386	bladder
FMS001	4	32743	fetal muscle
FSK001	8	127263	fetal skin
CLN001	4	28708	colon
REC001	3	28337	rectum
SPLc01	13	110573	spleen
FLG003	8	27360	fetal lung
THMc02	17	96791	thymus
NTU001	2	37055	neuronal cells
NTR001	2	34629	retinoic acid-induced neuronal cells
ABR006	365	108204	adult brain
FBR004	2	27560	fetal brain
FBR006	351	151893	fetal brain
ABR008	11420	145661	adult brain
FLS002	4	709733	fetal liver-spleen
IB2003	1108	201294	infant brain
ADP001	2	37287	cultured preadipocytes
FLV002	11	32865	fetal liver
PLA003	2	80877	placenta
FLV004	2	74491	fetal liver
ESO002	2	36840	esophagus
FSK002	4	72628	fetal skin
FMS002	7	40223	fetal muscle
FHR001	7	108446	fetal heart
FLS003	4	187791	fetal liver-spleen
HMP001	10	71425	macrophage
FLG004	1	41090	fetal lung
ABR016	57	45716	brain
BMD008	3	44770	bone marrow
LYN001	2	44025	lymph node
STM001	3	181899	bone marrow

SEQ ID NO: 54 was further analyzed for their presence in the public dbEST database and their tissue source. SEQ ID NO: 54 was found to be expressed in following tissues: Soares_total_fetus_Nb2HF8_9w, head_neck, kidney tumor, colon tumor RER+, Soares_fetal_heart_NbHH19W, head_neck, pooled germ cell tumors, kidney, subtracted, 2

pooled tumors (clear cell type), colon tumor RER+, malignant melanoma, metastatic to lymph node, LTI_NFL006_PL2, cervix carcinoma cell line, bone marrow cell line, melanotic melanoma, carcinoid, Pineal gland II.

The gene corresponding to SEQ ID NO: 54 was mapped to human chromosome 18p11.3 by BLAST analysis with human genome sequences.

By checking the Nuvelo proprietary database established from screening by hybridization, SEQ ID NO: 37 or 40 was found to be expressed in following human tissue/cell cDNA (see Table 45):

Table 45

Library Name	No. of Positive Clones	Total No. of Clones in the Library	Tissue Origin
ALV002	1	144402	adult liver
FBR006	1	151893	fetal brain
FKD002	1	33111	fetal kidney
FSK002	1	72628	fetal skin

SEQ ID NO: 37 or 40 was further analyzed for their presence in the public dbEST database and their tissue source. SEQ ID NO: 37 or 40 was found to be expressed in following tissues: Neuroblastoma cells.

The gene corresponding to SEQ ID NO: 37 or 40 was mapped to chromosome 12 by BLAST analysis with human genome sequences.

5.8 EXAMPLE 8

Expression Analysis of SEQ ID NO: 9

First strand human cDNA libraries from multiple tissues were screened with gene specific primers for SEQ ID NO: 9 [5'-CGATGCAGGAGAACCAGGAC- 3' (SEQ ID NO: 12 and 5'-CCTCAGGACCAGTGGGACC- 3' (SEQ ID NO: 13)]. The commercial panels (Clontech) screened were: Panel I (heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas), Panel II (Spleen, thymus, prostate, testis, ovary, small intestine, colon and adipocyte from a marathon ready cDNA library), immune panel (spleen, lymph node, thymus, tonsil, bone marrow, fetal liver, peripheral blood leukocyte) and a blood fraction panel (mononuclear, resting CD8+, resting CD4+, resting CD14+, resting CD19+, activated mononuclear cells, activated CD4+ and activated CD8+). PCR was performed for a total of 30 cycles using the following conditions: an initial denaturation at 94 °C for 3 min, followed by 5 cycles of 30 s at 94 °C, 30 sec at 68 °C and 1 min at 72 °C, followed by 5 cycles of 30 s at 94 °C, 30 sec at 64 °C and 1 min at 72 °C, followed by 20 cycles of 30 s at 94 °C, 30 sec at 60 °C and 1 min at 72 °C followed by

an extension of 10 min at 72 °C. The amplification product was detected by analysis on agarose gels stained with ethidium bromide. The SEQ ID NO: 9 was expressed in a human adipose tissue cDNA library.

5.9 EXAMPLE 9

Cellular Localization of SEQ ID NO: 10

SEQ ID NO: 9 specific primers corresponding to the translational start region and the carboxy-terminal region, excluding the stop codon of the SEQ ID NO: 9 sequence, were used [5'-TATAAGCTTATGAGGATCTGGTGGCTTCTG-3' (SEQ ID NO: 14) and 5'-AATCTCAGACGGGCTGCTGAACAGAAGG-3' (SEQ ID NO: 15)]. PCR amplification of the 883 nt product was performed using the following conditions; an initial denaturation at 94 °c for 3 min, followed by 5 cycles of 30 s at 94 °c, 30 sec at 66 °c and 1 min at 72 °c, followed by 5 cycles of 30 s at 94 °c, 30 sec at 62 °c and 1 min at 72 °c, followed by 20 cycles of 30 s at 94 °c, 30 sec at 58 °c and 1 min at 72 °c followed by an extension of 10 min at 72 °c. These primers generated a fragment of DNA corresponding to the entire coding region of the SEQ ID NO: 10, flanked by *HindIII* and *XhoI* sites. The PCR product was digested accordingly to generate overhang ends that were ligated to the *HindIII* and *XhoI* sites of PCDNA3.1/myc-His(+)A (Invitrogen). The resultant mammalian expression plasmid (AQL1/myc-His) allows for expression of the AQL1 coding sequence fused in-frame with the myc-6His epitope at the carboxy terminus.

The mammalian expression vector was transfected into COS-7 cells. Briefly, cells in a 10 cm dish with 8 ml of medium were incubated with 16 µl of Fugene-6 and 4 µg of DNA for 12 h. The medium was then replaced with serum-free DMEM and incubated for an additional 48 h prior to harvesting. After the conditioned medium was collected from transfected COS-7 cells, cells were washed twice with PBS and then scrapped from plates. Upon centrifugation, the cells were resuspended in PBS containing 0.5 µg/ml leupeptin, 0.7 µg/ml pepstatin, and 0.2 µg/ml aprotinin. After a brief sonication, the cytosolic fraction was separated from the insoluble membrane fraction by centrifugation. Purification of proteins from the cytosolic and from the media took place at 4 C in the presence of 100 µl of Ni-NTA resin (Qiagen). The resin was washed twice with 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, and 5 mM imidazole

To determine the cellular localization of the AQL1/myc—His tagged protein, Western blot analysis was performed on cytosolic, membrane, and medium fractions using an anti-myc antibody. AQL1/myc—His tagged protein was detected primarily in the medium (85%), but some protein was also detected in the cytosolic (10%) and membrane (5%) fractions. The

predicted molecular mass of the tagged AQL1/myc—His tagged protein is 38 kDa. However, the approximate 44 kDa electrophoretic mobility suggests that AQL1/myc—His tagged protein is post-translationaly modified.

5.10 EXAMPLE 10

Chromosomal Localization of SEQ ID NO: 10

To determine the chromosomal localization of SEQ ID NO: 10, gene specific PCR primers [5'-AAGCCTGGTCCCAAAGGAGA-3' (SEQ ID NO: 15) and 5'-GGTGTGGCGGATTTTTAAACTCT-3' (SEQ ID NO: 16)] were screened against the NIGMS human/rodent somatic cell hybrid mapping panel #2. PCR amplification of the 423 nt product was performed using the following conditions; an initial denaturation at 94 °C for 3 min, followed by 5 cycles of 30 s at 94 °C, 30 sec at 68 °C and 1 min at 72 °C, followed by 5 cycles of 30 s at 94 °C, 30 sec at 64 °C and 1 min at 72 °C, followed by 20 cycles of 30 s at 94 °C, 30 sec at 60 °C and 1 min at 72 °C followed by an extension of 10 min at 72 °C. All products were separated by 3% agarose gel electrophoresis and visualized via ethidium bromide staining. SEQ ID NO: 10 was mapped to chromosome 13.

5.11 EXAMPLE 11

Multiplex Analysis of Phosphorylation Status of Different Signaling Molecules After Treatment with AQL1 Polypeptide

Protein phosphorylation is one of the most common post-translation modifications involved in transmitting extracellular signal to intracellular target molecules. Phosphorylation of intracellular protein is regulated by proteins called kinases. Measuring protein phosphorylation provides a tool for predicting the activity of a protein. An increase or decrease of intracellular protein phosphorylation after treatment of a cell type with C1q domain-containing protein could be an indication of a potential function of C1q domain-containing protein in this cell type. The assay is carried out in a Bio-Plex (BioRad) and the multiplex phosphoprotein assay measures levels of phospho-JNK, phospho-p38MAPK, phospho-erk, phospho-stat3, phospho-IkBalpha, phospho-akt, total tyrosine phosphorylation and phospho-EGF.

5.12 EXAMPLE 12

Calcium Mobilization Assay

Many extracellular signals to intracellular targets are mediated by increases in free calcium levels in the cytoplasm. Calcium mobilization from intracellular stores can be detected

in many cell types by loading the cells with a Ca^{2+} sensitive indicator such as fura-2- AM. The increase in fluorescence is detected by a fluorescence plate reader. Cells will be incubated in media containing 5 μM Fura-2 AM, 5 μM Pluronic F-127 for 30 min. After the addition of C1q domain-containing protein the Fura-2 intensity will be monitored approximately every 20 sec by a fluorescent plate reader (Molecular Dynamics) and compared to the intensity of cells with basal calcium levels.

5.13 EXAMPLE 13

Fatty Acid Oxidation Assay

The oxidation of palmitate or oleate in culture C2Cl2 skeletal muscle cells (ATCC; CRL-1772) upon exposure to AQL1 protein is measured according to published procedures (Barger et al., J. Clin. Invest. 105:1723-1730 (2000)). In summary, nearly confluent C2C12 myocytes are kept in differentiation medium (DMEM, 2.5% horse serum) for 7 days, at which time formation of myotubes is maximal. [1-^{14}C]oleic acid (1 μCi/ml) is added to the cells and incubated for 90 minutes at 37°C in the absence/presence of C1q domain-containing protein. In some of the assays a proteolytically cleaved C1q domain-containing protein (cleaved between lysine 190-glycine 191) may be employed. During the experiment the C2C12 cells are incubated in a closed system containing Whatman paper to collect the $^{14}CO_2$ gas released during fatty acid oxidation. After the incubation the Whatman paper is removed and the amount of ^{14}C radioactivity is determined by liquid scintillation counting.

5.14 EXAMPLE 14

Macrophage Phagocytosis Assay

Human macrophages are incubated in the presence/absence of C1q domain-containing protein for 24 hours at 37°C in 96-well plates. Fluobrite fluorescent-microspheres (0.75G; Polyscience, Warrington, PA) are added to each well, followed by one hour incubation at 37°C. Nonadherent latex beads are removed by gentle washing and the cells are incubated for an additional 30 minutes to complete phagocytosis. The cells are harvested by short-time treatment with EDTA and trypsin and washed vigorously three times with PBS to remove noningested beads. The amount of ingested beads will be measured with a FACScan.

5.15 EXAMPLE 15

Expression Study Using SEQ ID NO: 1-3, 6, 9, 12, 15-17, 20-22, 24, 27-28, 31, 34-36, 38, 43-45, or 52-54

The expression of SEQ ID NO: 1-3, 6, 9, 12, 15-17, 20-22, 24, 27-28, 31, 34-36, 38, 43-45, or 52-54 in various tissues is analyzed using a semi-quantitative polymerase chain reaction-based technique. Human cDNA libraries are used as sources of expressed genes from tissues of interest (adult bladder, adult brain, adult heart, adult kidney, adult lymph node, adult liver, adult lung, adult ovary, adult placenta, adult rectum, adult spleen, adult testis, bone marrow, thymus, thyroid gland, fetal kidney, fetal liver, fetal liver-spleen, fetal skin, fetal brain, fetal leukocyte and macrophage). Gene-specific primers are used to amplify portions of SEQ ID NO: 1-3, 6, 9, 12, 15-17, 20-22, 24, 27-28, 31, 34-36, 38, 43-45, or 52-54 sequences from the samples. Amplified products are separated on an agarose gel, transferred and chemically linked to a nylon filter. The filter is then hybridized with a radioactively labeled (^{33}P-dCTP) double-stranded probe generated from SEQ ID NO: 1-3, 6, 9, 12, 15-17, 20-22, 24, 27-28, 31, 34-36, 38, 43-45, or 52-54 using a Klenow polymerase, random-prime method. The filters are washed (high stringency) and used to expose a phosphorimaging screen for several hours. Bands indicate the presence of cDNA including SEQ ID NO: 1-3, 6, 9, 12, 15-17, 20-22, 24, 27-28, 31, 34-36, 38, 43-45, or 52-54 sequences in a specific library, and thus mRNA expression in the corresponding cell type or tissue.